CuttingEEG26-FFM | 23 Sep 2026 | CoBIC

Workshop

Building a 1,000-Song Naturalistic EEG Dataset: Practical Lessons from ManyMusic🎶

Slides PDF

Seung-Goo Kim

Research Group Neurocognition of Music and Language
Max Planck Institute for Empirical Aesthetics, Frankfurt am Main, Germany

Workshop plan

Building a 1,000-Song Naturalistic EEG Dataset

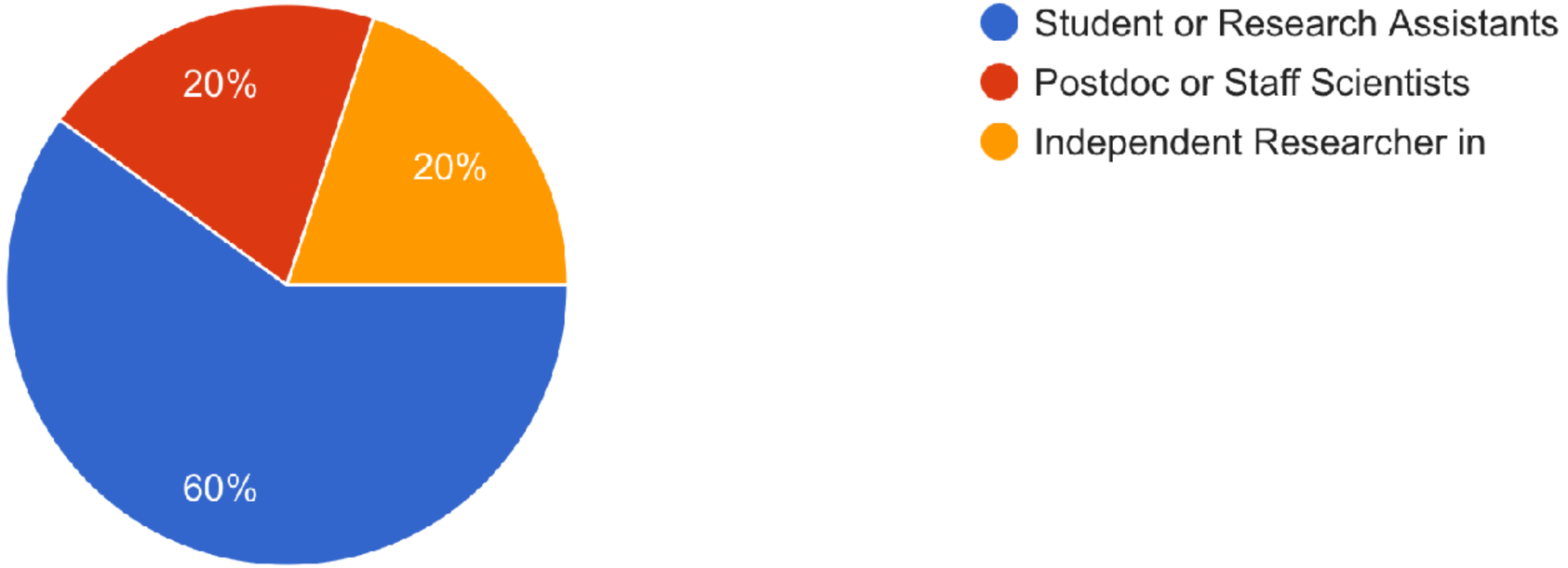
Part	From	To	Contents	Dur
1	09:00	09:45	**Introduction**	45m
2	10:00	10:45	**Analysis**	45m
3	11:00	11:45	**Design**	45m
4	12:00	12:45	**Management**	45m

Pre-survey result

N=5

What is your role?

5 responses

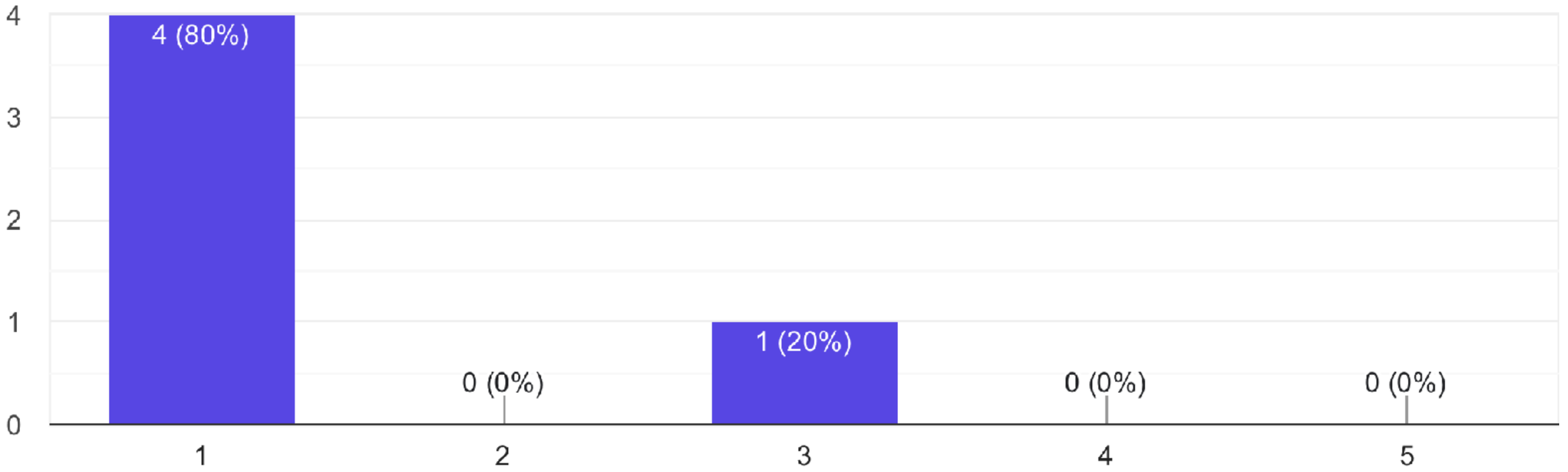


Pre-survey result

N=5

How familiar/experienced are you with the naturalistic neuroimaging datasets? (e.g., THINGS, Natural Scence Dataset, Narrative Dataset)

5 responses

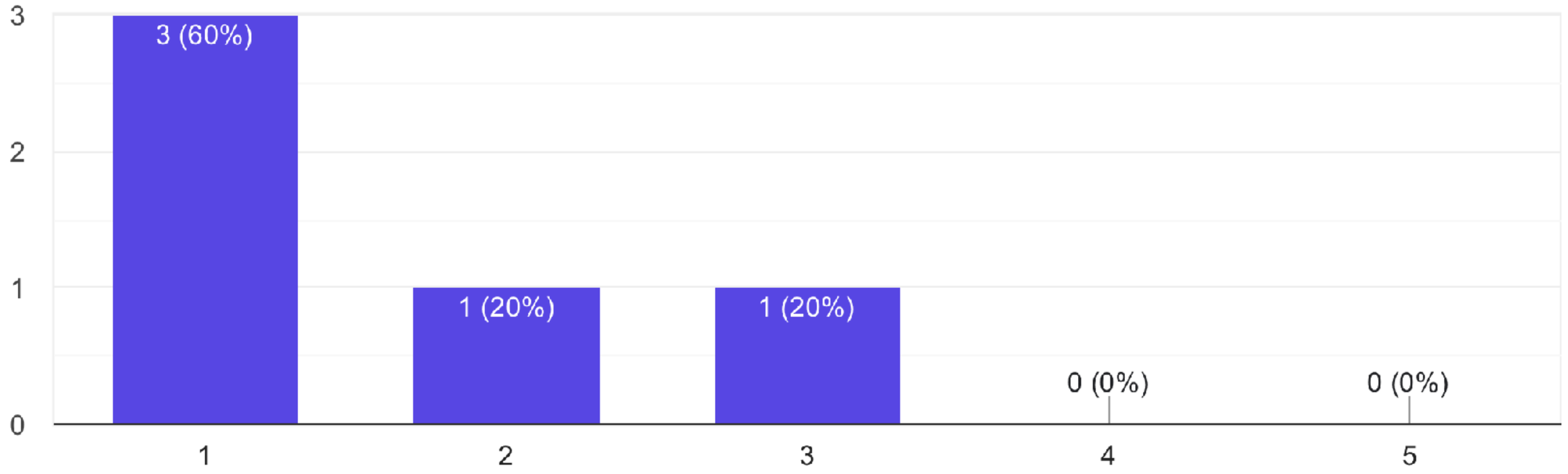


Pre-survey result

N=5

How familiar/experienced are you with the encoding analysis (i.e., ridge regression with the FIR models)?

5 responses

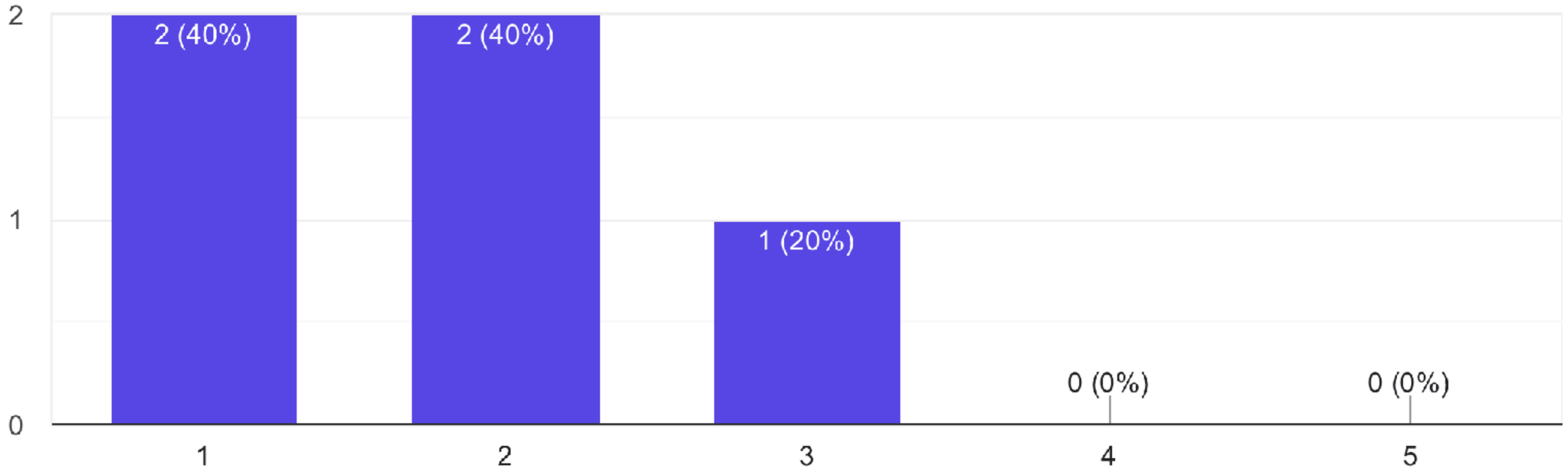


Pre-survey result

N=5

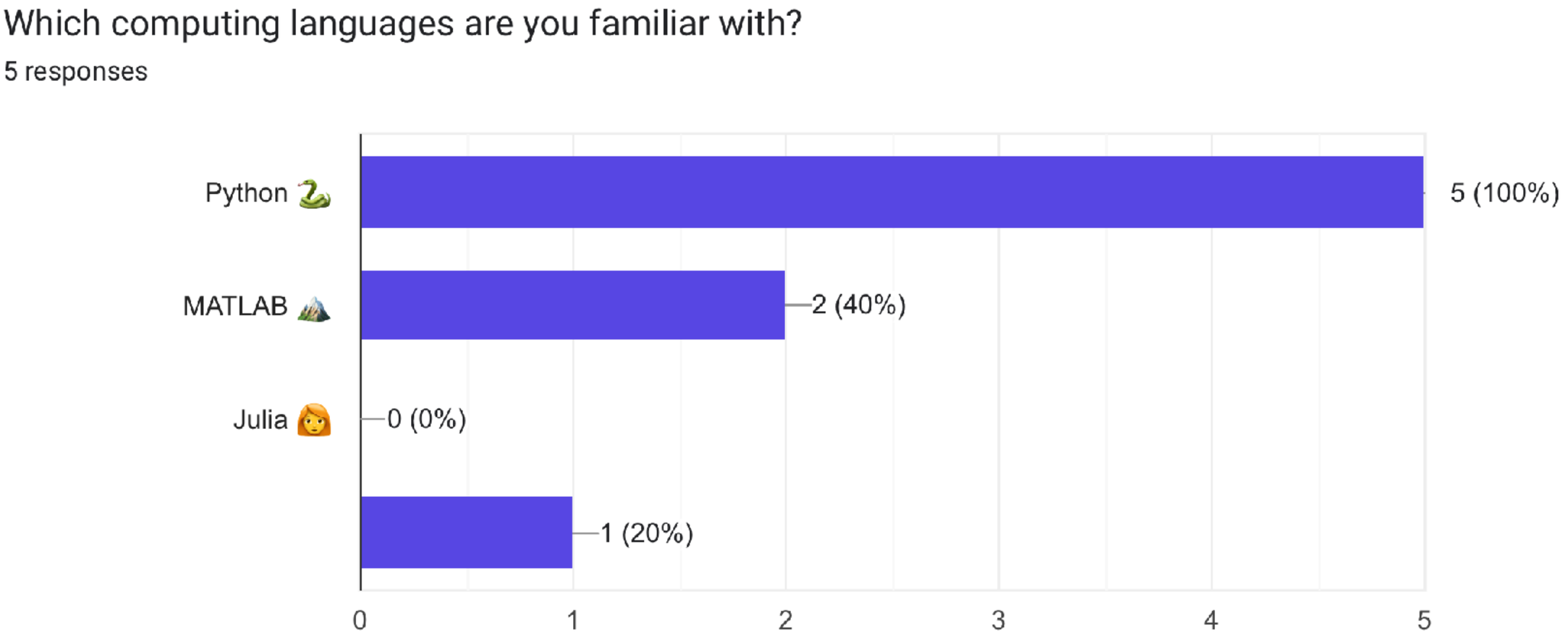
How familiar/experienced are you with the stimulus feature modeling (from sensory to semantic features)?

5 responses



Pre-survey result

N=5



Before we start…

Self-introduction

Open Recent Project

Sign In

def short_intro(person):

```
def short_intro(person):
    print(f"NAME={person.name}, TEAM={person.teams[0]}, ROLE=
    {person.teams[0].roles[0]}, PASSION_PROJECT={person.projects
    [0].name}")
for person in EVERYONE:
    short_intro(person)

```

-- NORMAL -- 5:1 Python

Workshop plan

Building a 1,000-Song Naturalistic EEG Dataset

Part	From	To	Contents	Dur
1	09:00	09:45	**Introduction**	45m
2	10:00	10:45	**Analysis**	45m
3	11:00	11:45	**Design**	45m
4	12:00	12:45	**Management**	45m

Naturalistic Stimuli for Music Research

🤔: But what do you mean by "naturalistic"?



https://voxpopulisphere.com/2015/11/01/audio-what-ancient-greek-music-actually-sounded-like/

How does music evoke emotions *via* the brain?

Neuroscientific view

Music





Brain



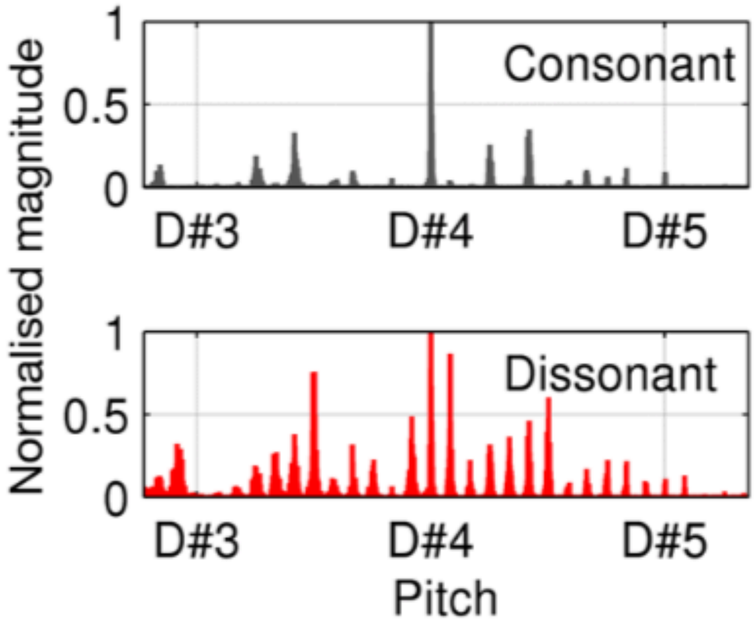


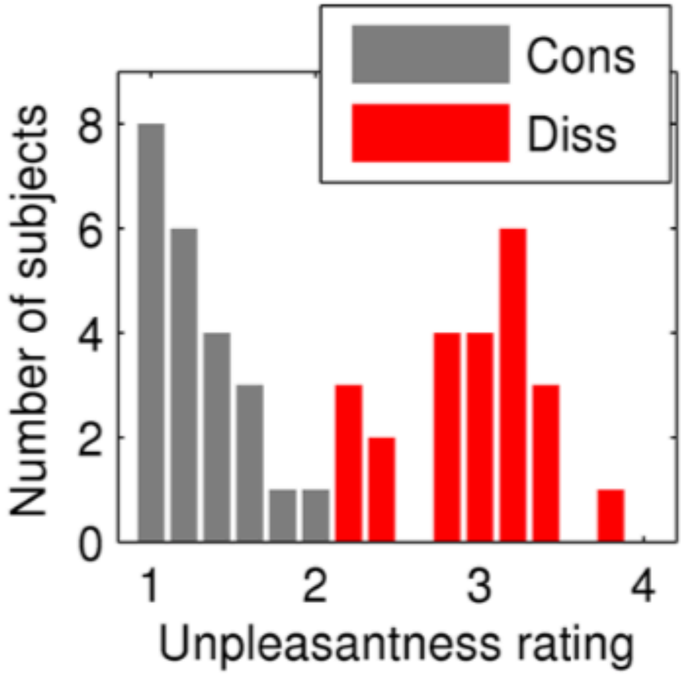
Behavior



Traditional approach: music vs. "non-music"

Music  →  Behavior



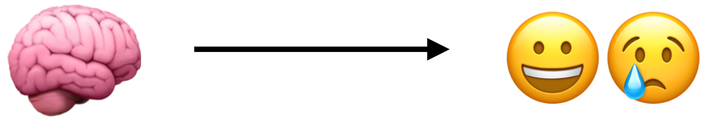


“Dissonant” stimuli were rated more unpleasant.

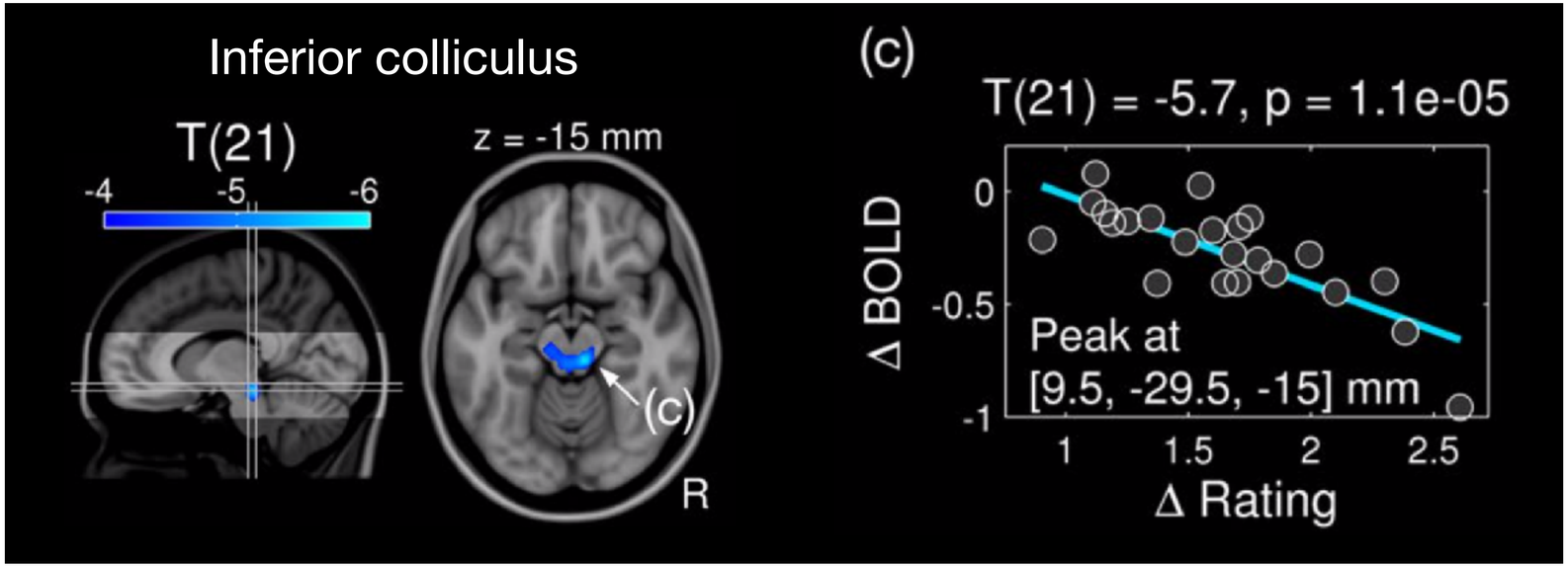
stimuli per condition = 20, stimulus duration = 30 s, # subjects = 23 (25.9 yo; 13 F)

Kim et al., 2017, *Sci Rep.*

Traditional approach: music vs. "non-music"



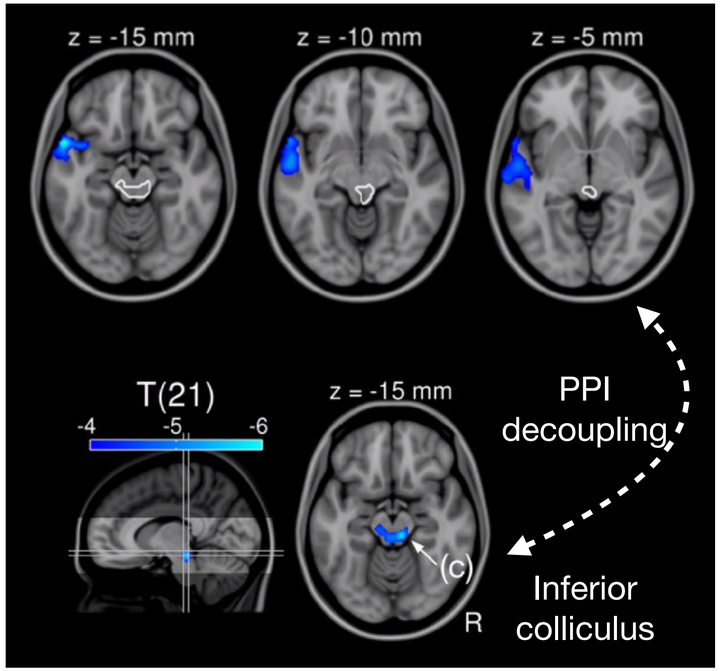
$$\Delta BOLD = \beta_0 + \beta_1 \Delta Rating + error$$



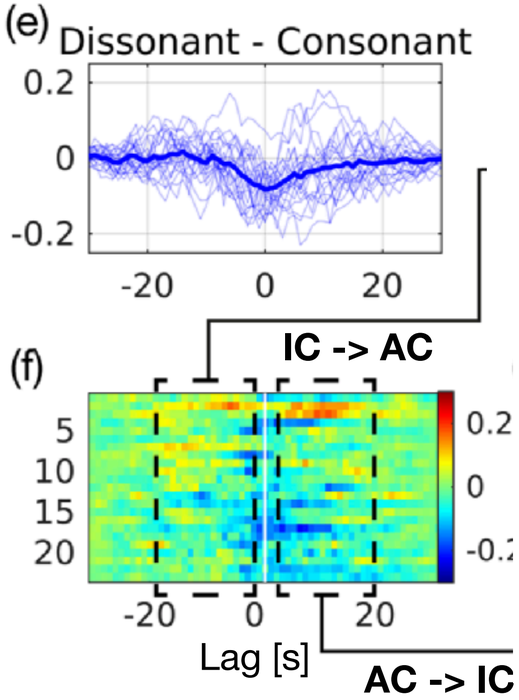
BOLD activation in the inferior colliculus (IC) decreased more in individuals who rated dissonant versions worse.

Traditional approach: music vs. "non-music"

Psycho-Physiological Interaction (PPI)



Cross-correlation



The IC decoupled more from the left auditory cortex (top-down) in individuals who rated dissonant versions worse.

Limitations of controlled stimuli

Ceteris paribus... and the subject is not listening.

Disruptive manipulations: validity, generalisability, & specificity?



🤷: **But what else can we do?**

Alternative approach: "naturalistic neuroimaging"

Embracing complexity for ecological validity

Gibson, 1978, The Ecological Approach to the Visual Perception of Pictures, *Leonardo.*

Rieke, 1995, Naturalistic Stimuli Increase the Rate and Efficiency of Information Transmission by Primary Auditory Afferents, *Proceedings of the Royal Society B.*

Theunissen et al., 2000, Spectral-Temporal Receptive Fields of Nonlinear Auditory Neurons Obtained Using Natural Sounds, *Journal of Neuroscience*.

Hasson et al., 2004, Intersubject Synchronization of Cortical Activity During Natural Vision, *Science*.

Kay et al., 2008, Identifying Natural Images From Human Brain Activity, *Nature*.

Mesgarani et al., 2014, Phonetic Feature Encoding in Human Superior Temporal Gyrus, *Science*.

Huth et al., 2016, Natural Speech Reveals the Semantic Maps That Tile Human Cerebral Cortex, *Nature*.

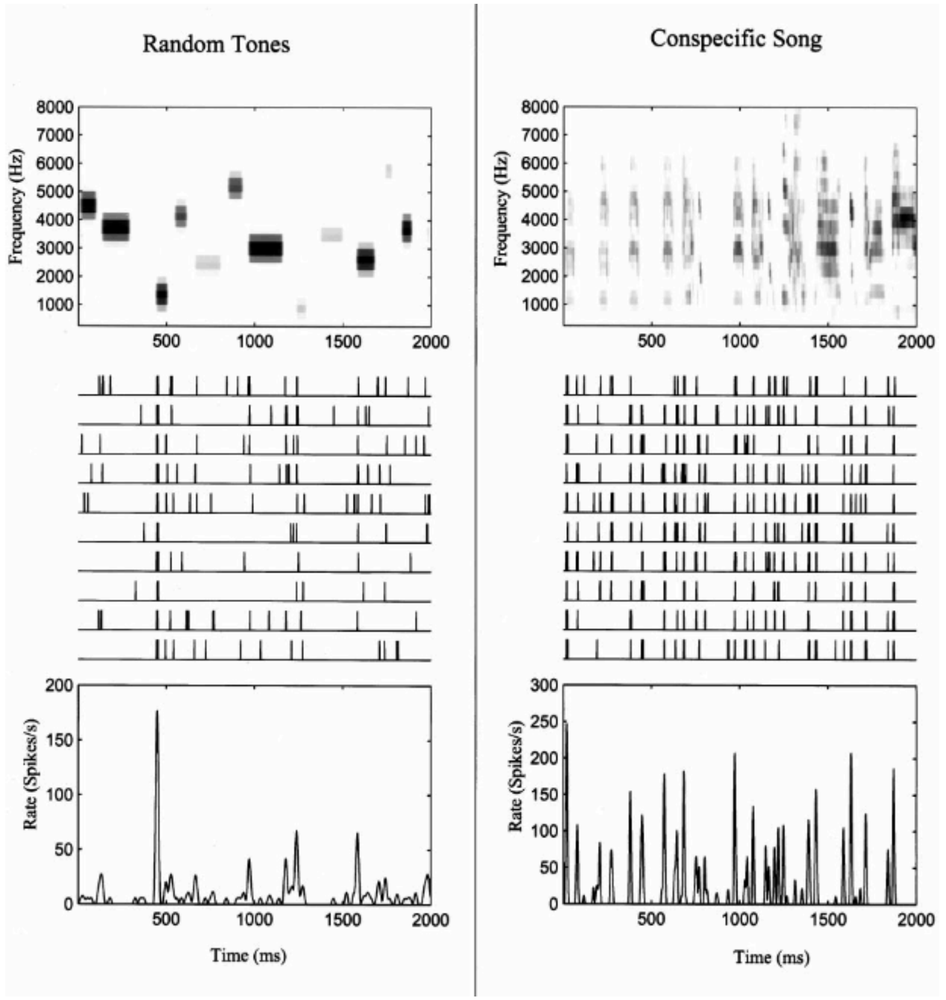
Hamilton & Huth, 2020, The Revolution Will Not Be Controlled: Natural Stimuli in Speech Neuroscience, *Language, Cognition and Neuroscience*.

1980 1990 2000 2010 2020

Receptive field mapping of a conspecific calls

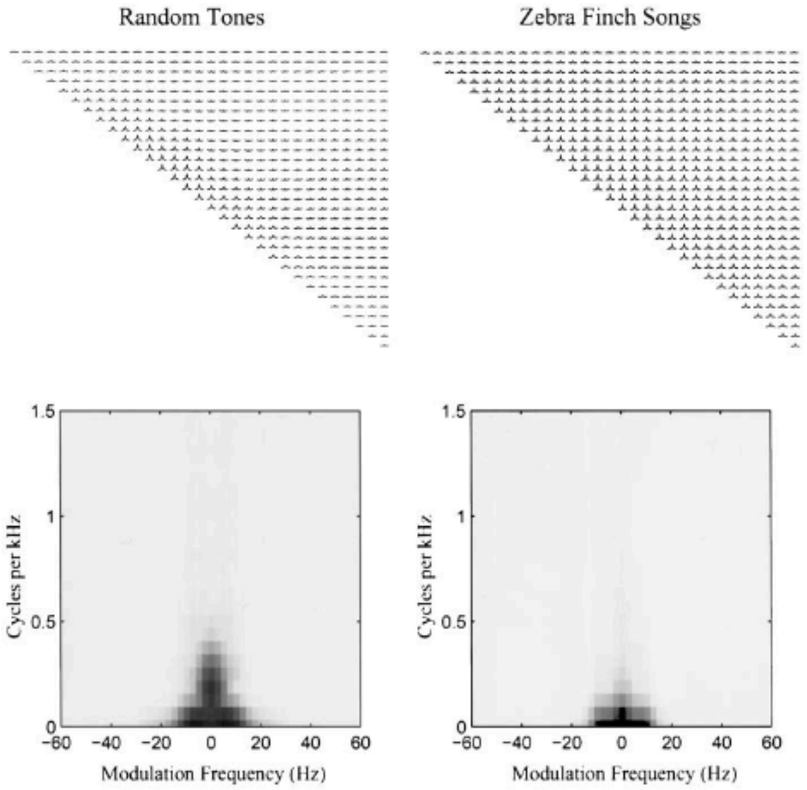
Invasive recording of local field potential of neurons

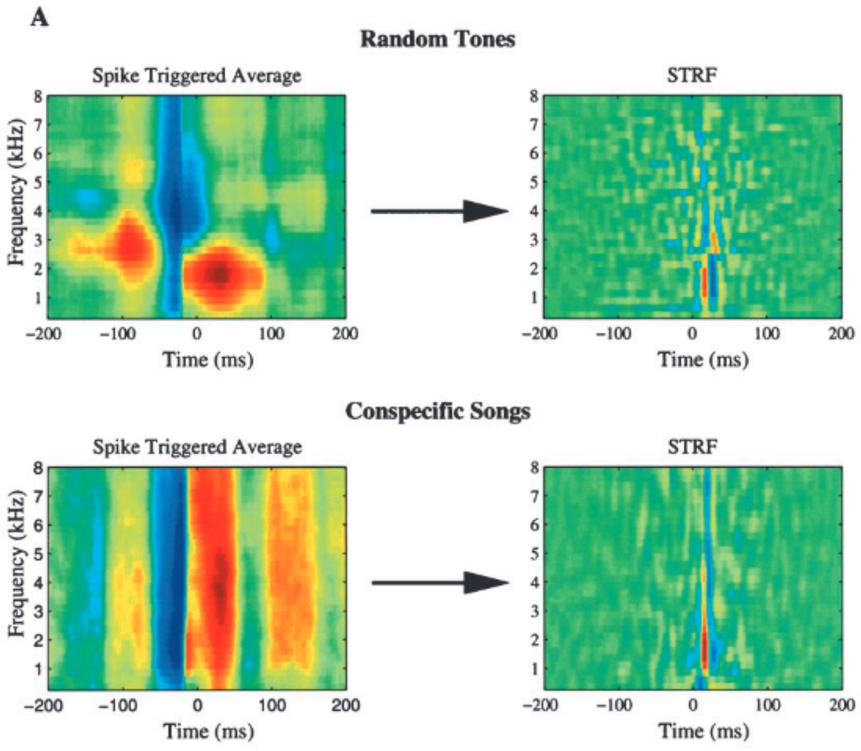


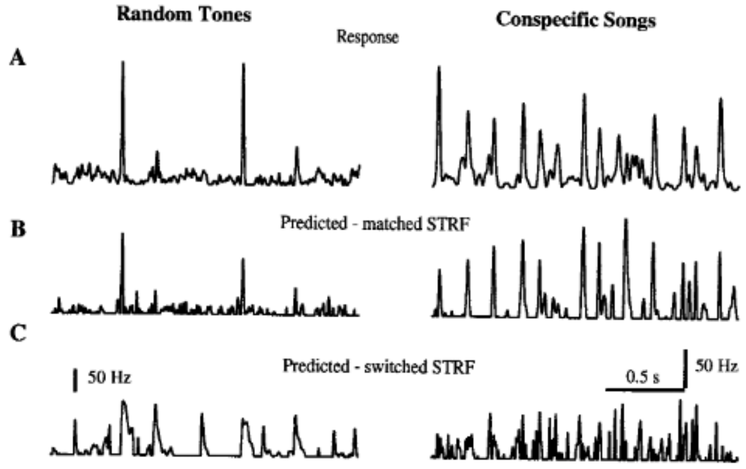


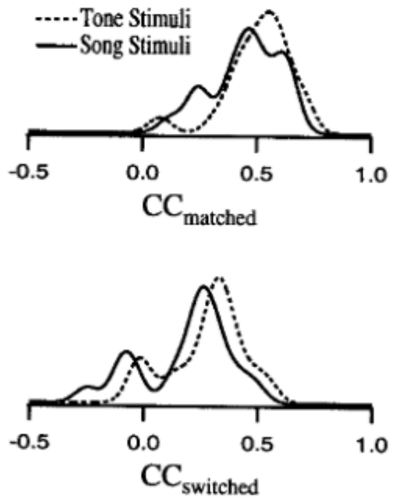
Receptive field mapping of a conspecific calls

Invasive recording of local field potential of neurons



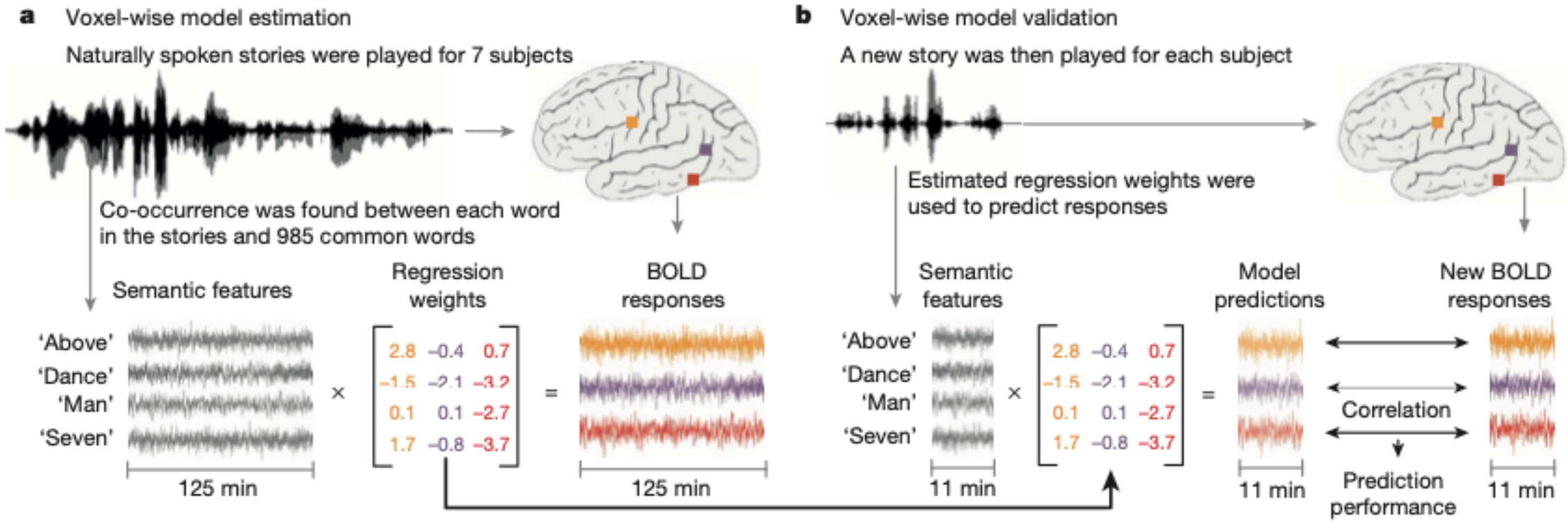


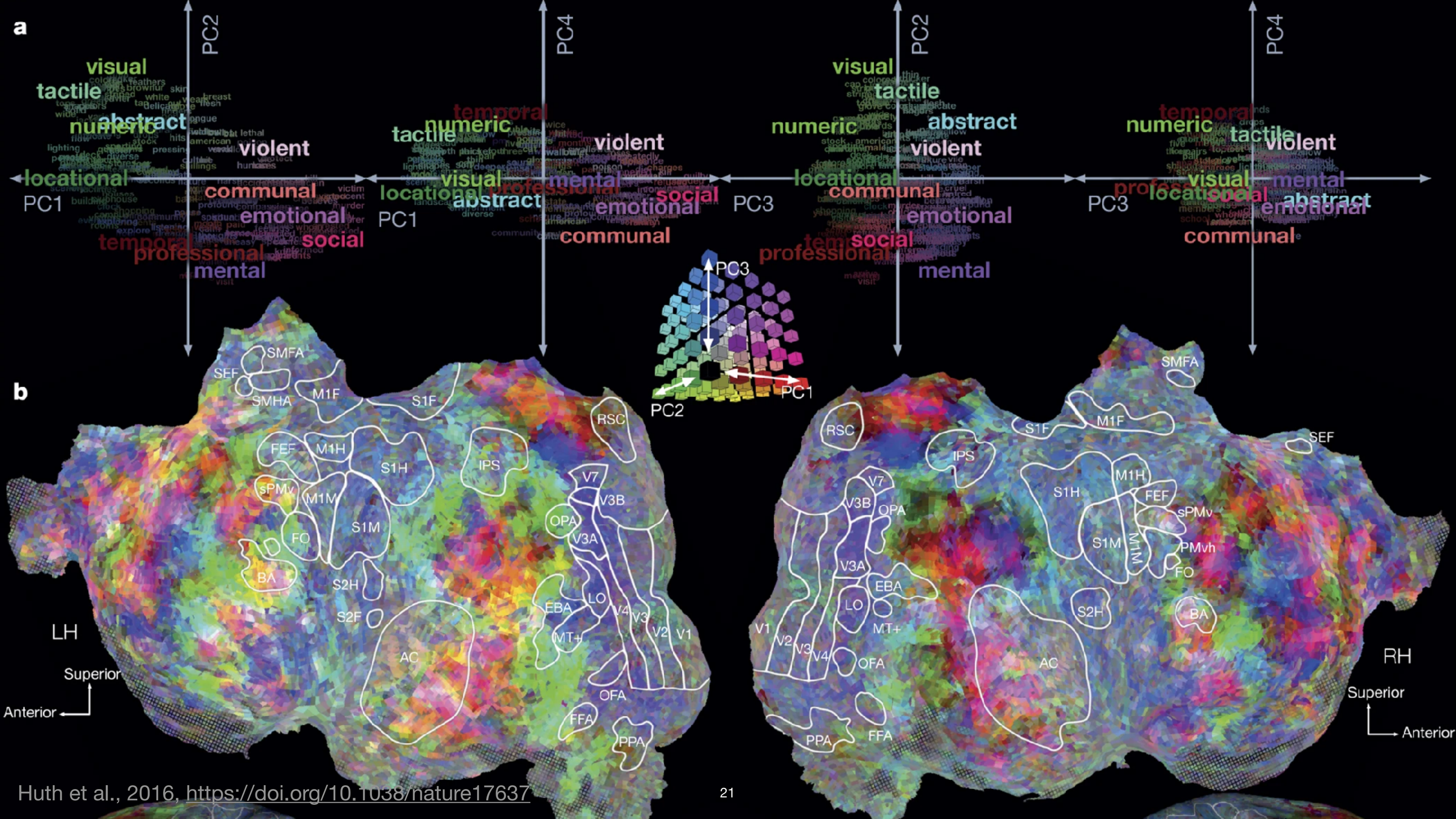




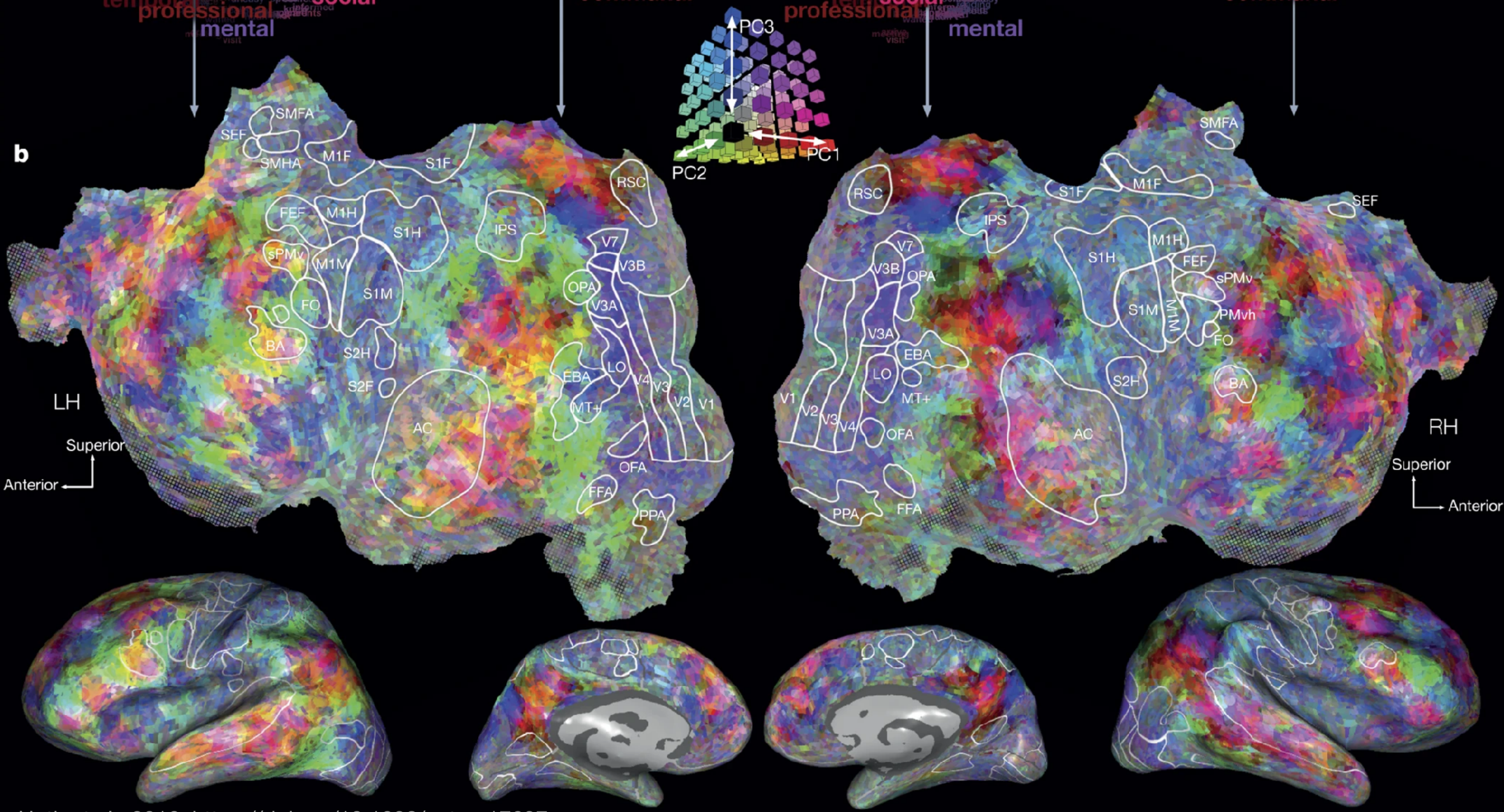
Semantic representation in the whole cortex

Non-invasive scanning of functional MRI (local paramagnetism)



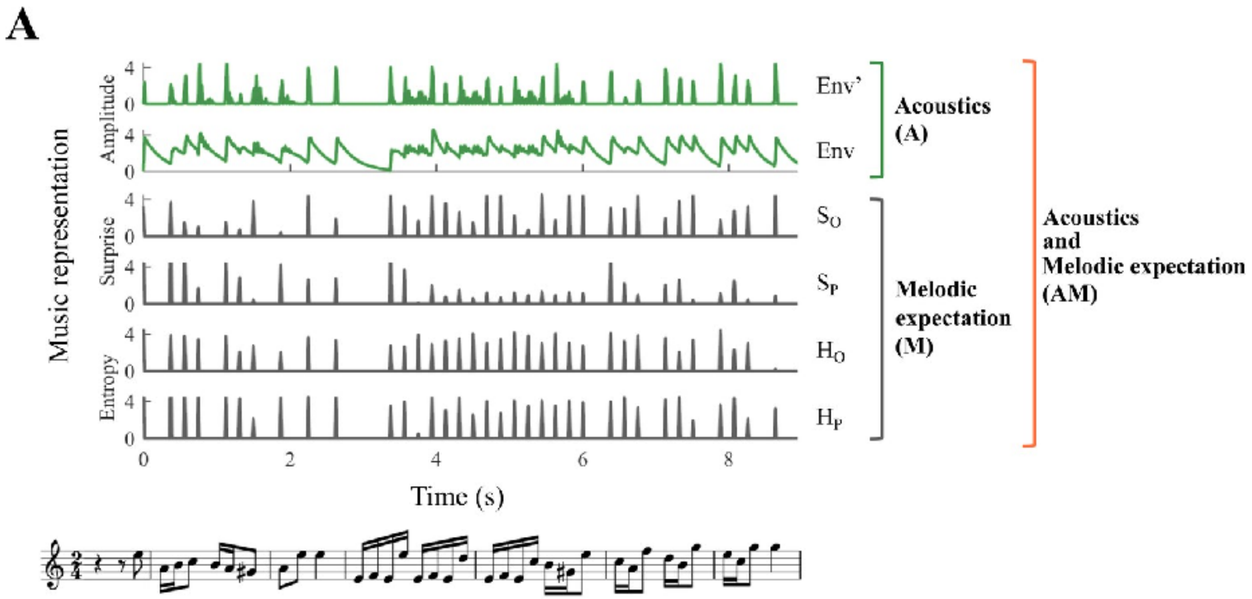


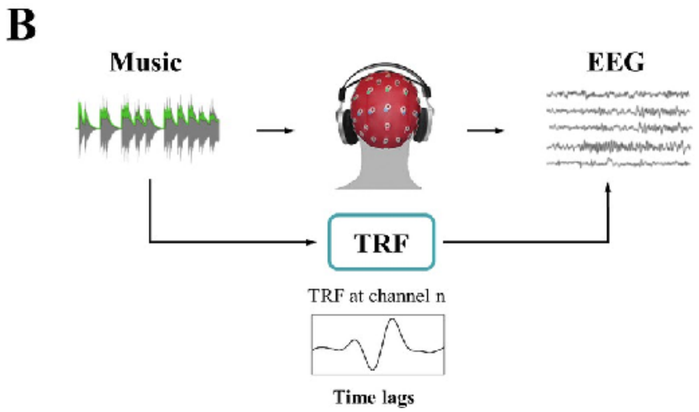
Huth et al., 2016, https://doi.org/10.1038/nature17637

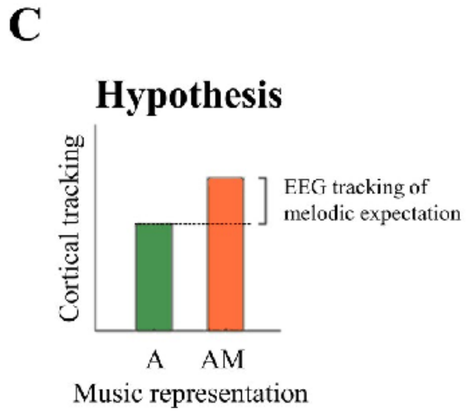

professional
mental
PC3
PC2
PC1
b
SMFA
SEF
SMHA
M1F
S1F
RSC
FEF
M1H
S1H
IPS
V7
sPMv
M1M
V3B
OPA
S1M
V3A
FO
BA
S2H
LO
EBA
V4
V3
V2
V1
MT+
S2F
AC
OFA
FFA
PPA
LH
Superior
Anterior
RH
PMvh

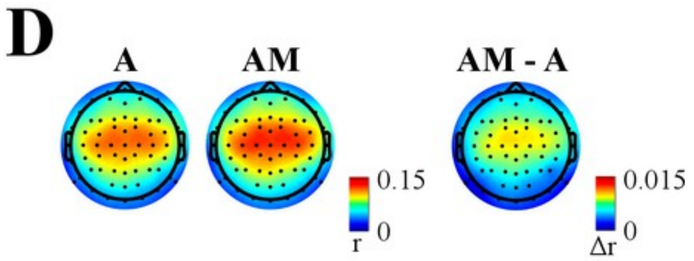
Temporal response function of melodic surprisal

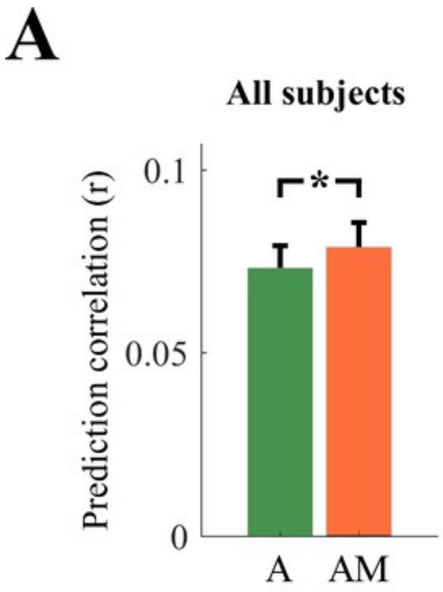
Non-invasive recording of EEG (scalp potential)

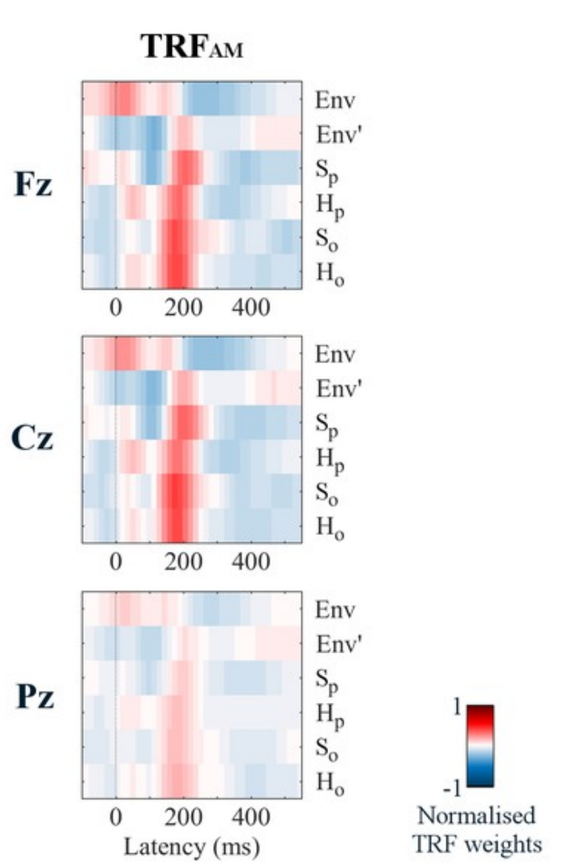












Any questions so far?

Example: abstraction gradient of musical information (in prep)

What in the music evoke emotions *via* the brain?

Music



Brain

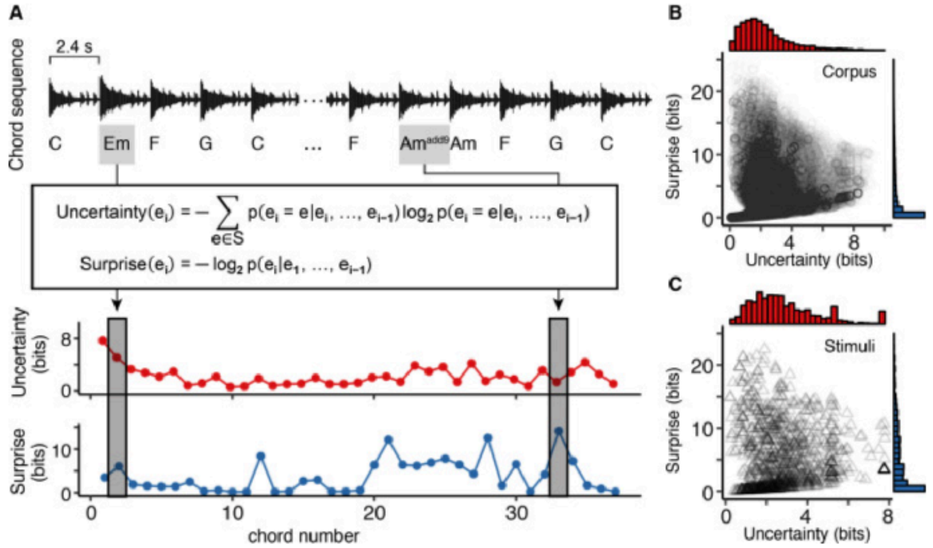


Emotion

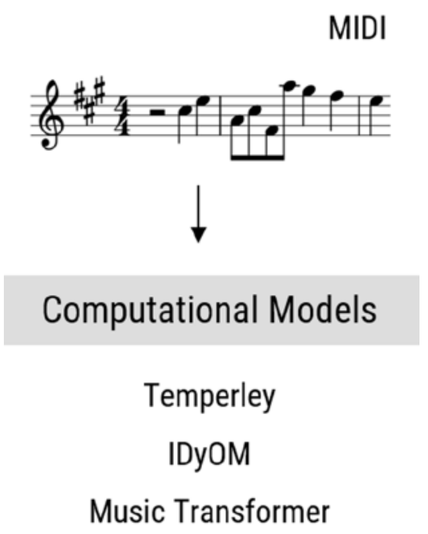


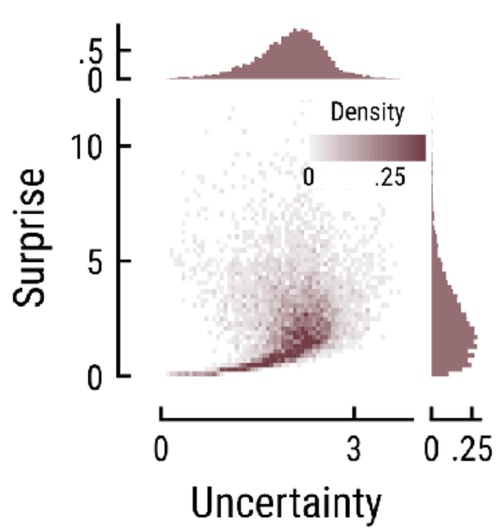
Symbol-domain models

Modeling on MIDI notes: **n-gram modeling** (e.g., PPM, HMM),
deep neural networks (e.g., Melody RNN, MusicRNN, Music Transformer)



Cheung et al., 2019, *Current Biology.*



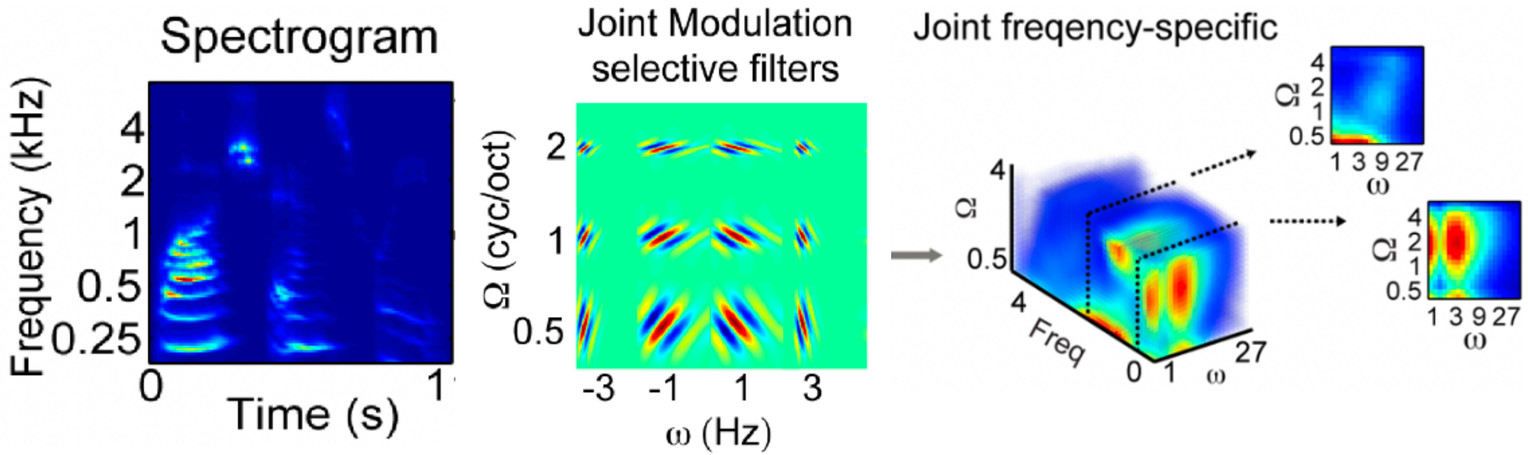


Kern et al., 2022, *eLife*.

Limited to transcription and tokenisation

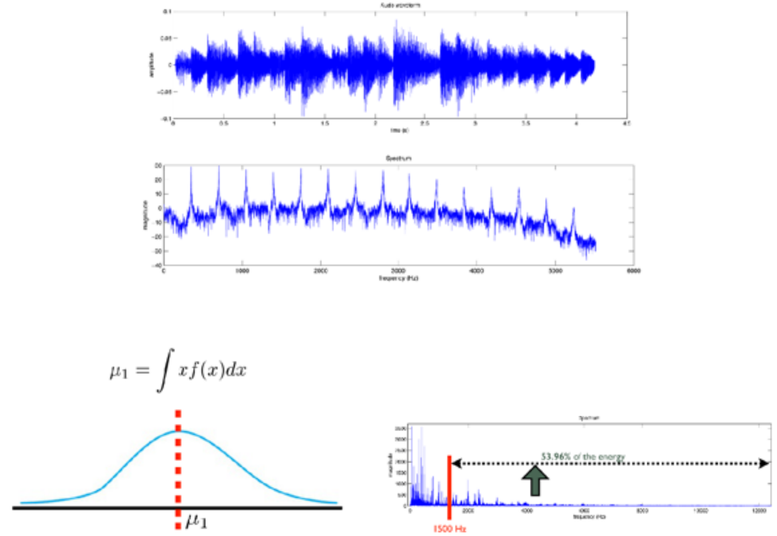
Audio-domain models

Auditory models: simulated activity of peripheral and central auditory neurons



Santoro et al., 2014, *PLOS Comp Biol*.

(Pre-DNN) Music information retrieval (MIR) models: acoustic features relevant to music database



Lartillot, 2017, *MIRtoolbox User's Manual.*

Mostly cochlear and primary auditory cortex activity.

Semantic audio models

VGGish (Hershey et al., 2017👴, ICASSP)

- VGG (Visual Geometry Group, Oxford University): a CNN for image classification
- VGG-ish: a VGG-like CNN for audio classification
- Data: YouTube-100M (100M video data labeled with ≥1 "topics" from 30K labels)
- Loss: audio-based label (updated) vs. video-based annotations (fixed)

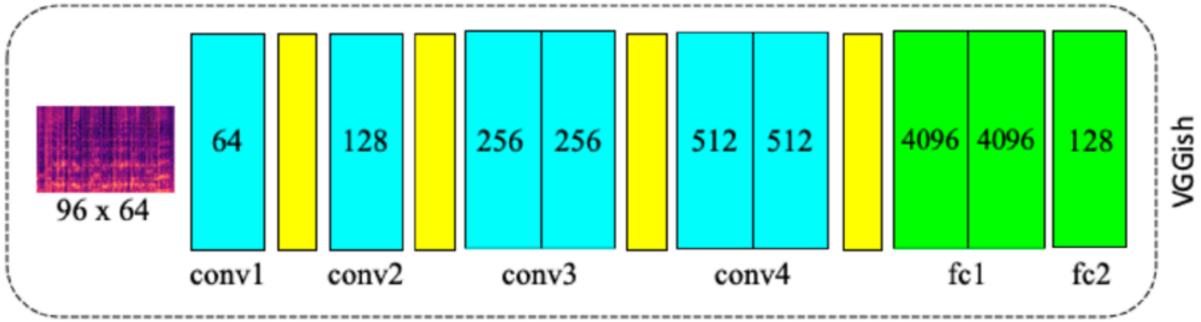
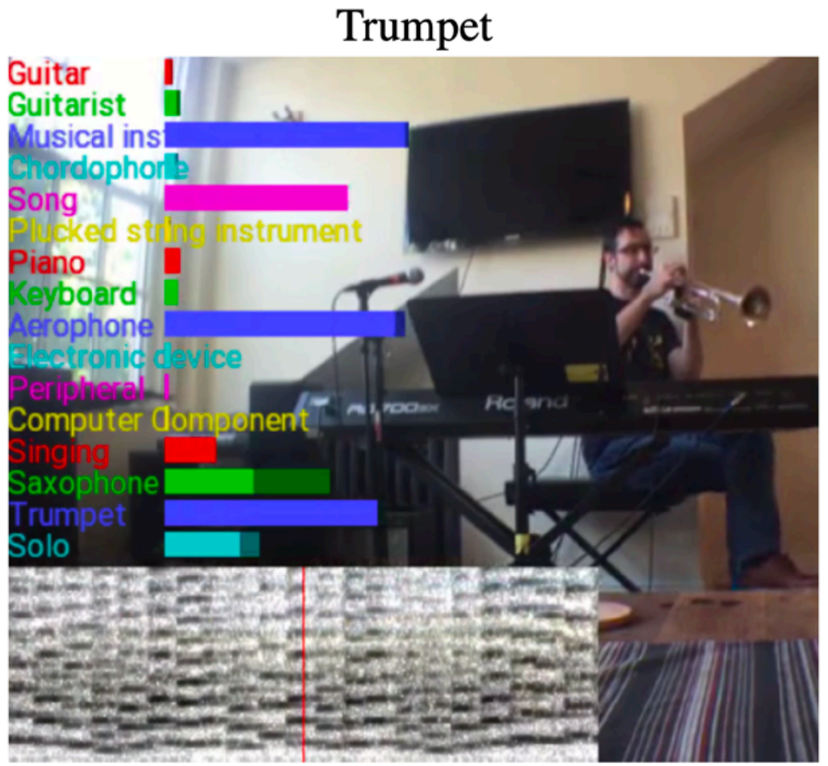


Table 1: Example labels from the 30K set.

Label prior	Example Labels
0.1 . . . 0.2	Song, Music, Game, Sports, Performance
0.01 . . . 0.1	Singing, Car, Chordophone, Speech
$\sim 10^{-5}$	Custom Motorcycle, Retaining Wall
$\sim 10^{-6}$	Cormorant, Lecturer

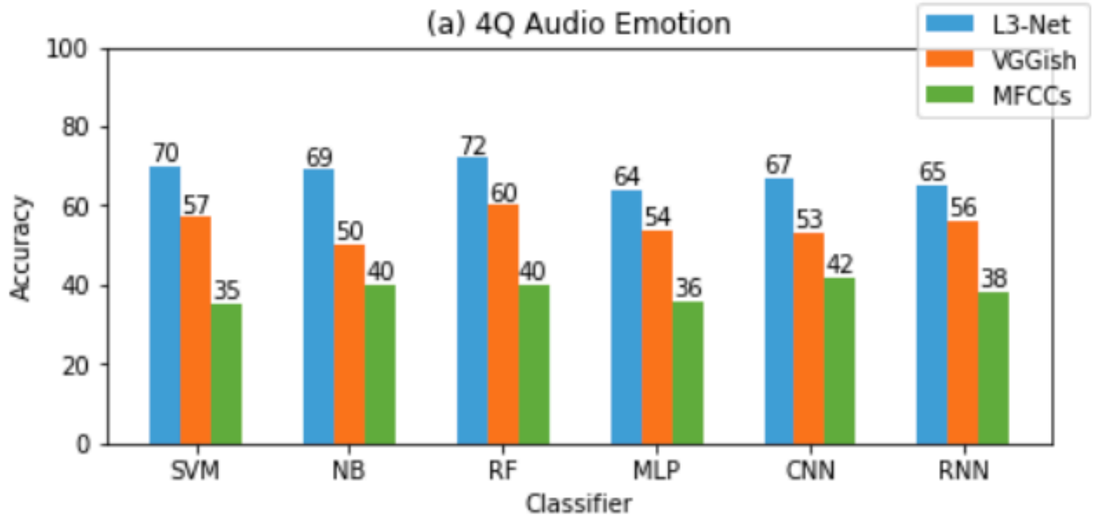


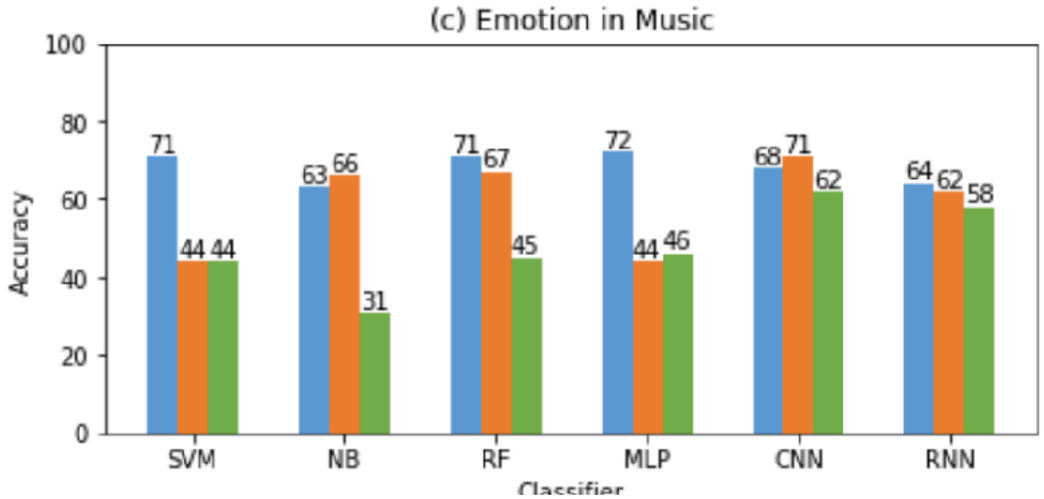
A possible mid/high-level representation of sounds

CNN embedding for music emotion *recognition*

Potentially mid/high-level representation of music signal

Classifiers (SVM, NB, RF, ...) based on MFCC/VGGish/L3 embeddings to predict *recognised* emotional categories





Koh & Dubnov, 2021, *ACA*

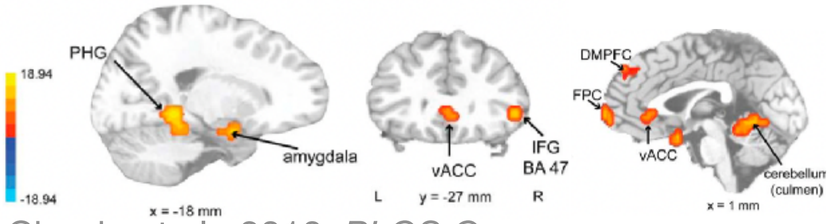
Deep audio semantic models carry more information related to expressed emotions than a traditional audio descriptor.

Example: Expressive vs. Mechanical

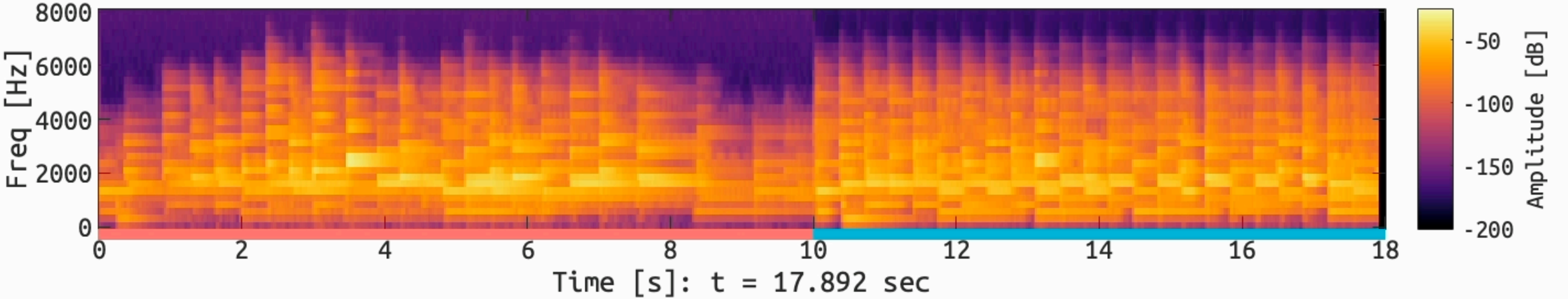
Chapin et al., 2010, PLOS ONE.

[EXPRESSIVE]: "Frédéric Chopin's Etude in E major, Op.10, No. 3 was performed by an undergraduate piano major (female, 22 years old) on a Kawai CA 950 digital piano"

Expressive > Mechanical



Chapin et al., 2010, *PLOS One*.

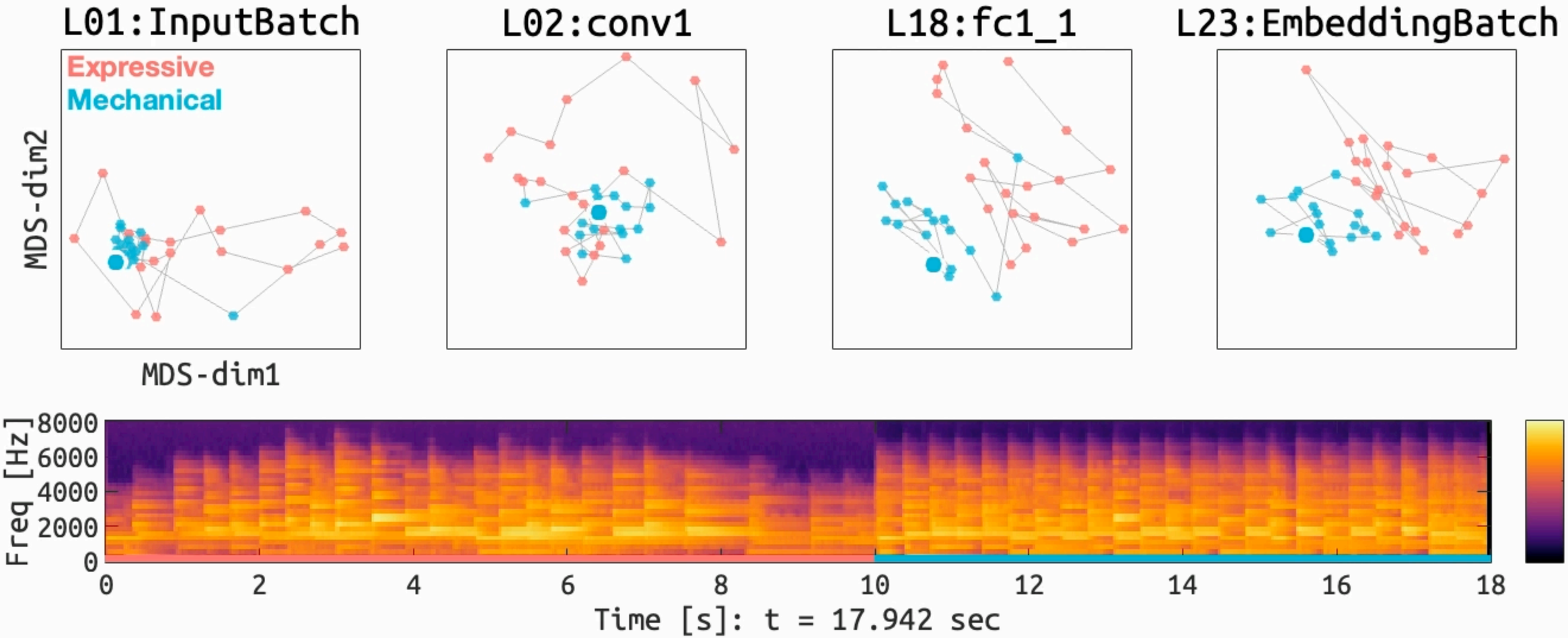


[MECHANICAL]: "The MIDI (Musical Instrument Digital Interface) onset velocity (key pressure) of each note (correlating with sound level) was set to 64 (range 0–128), and pedal information was eliminated."

Audio source: https://doi.org/10.1371/journal.pone.0013812

Example: Expressive vs. Mechanical

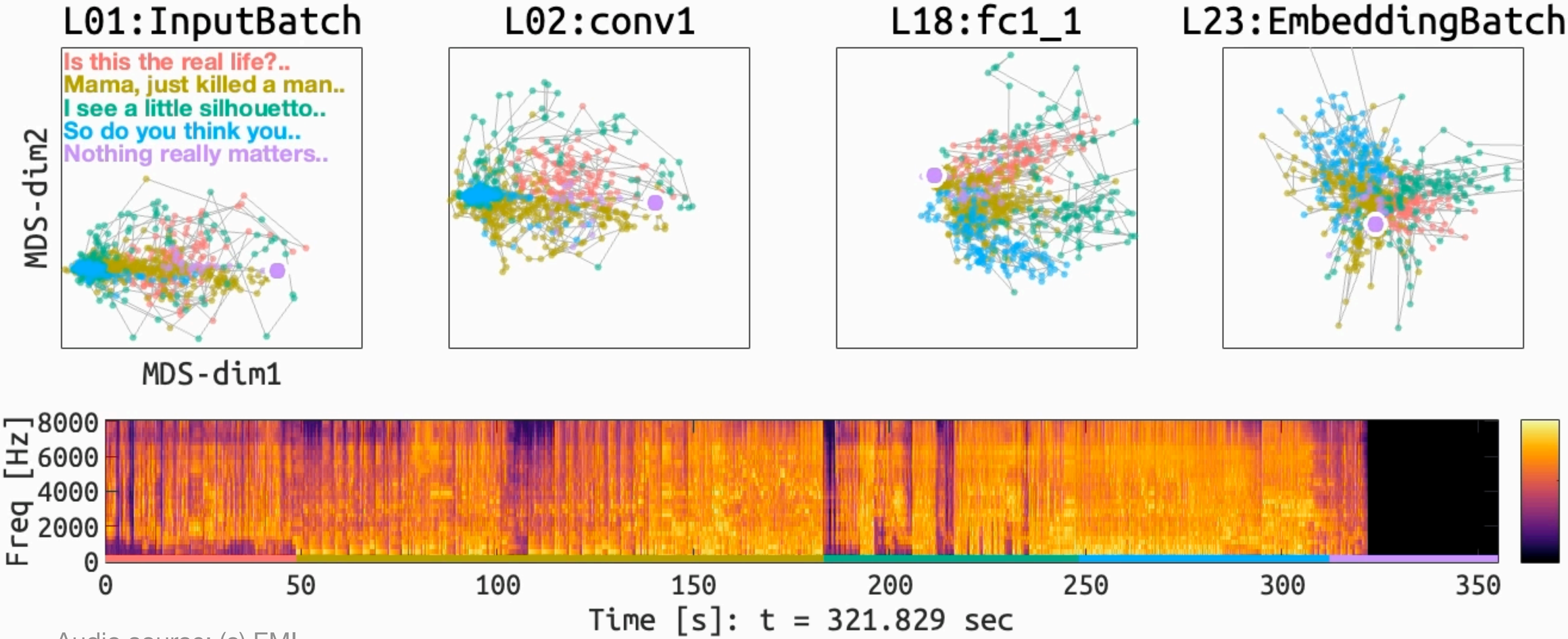
VGGish representations at various layers (from 1 to 23)



Audio source: https://doi.org/10.1371/journal.pone.0013812

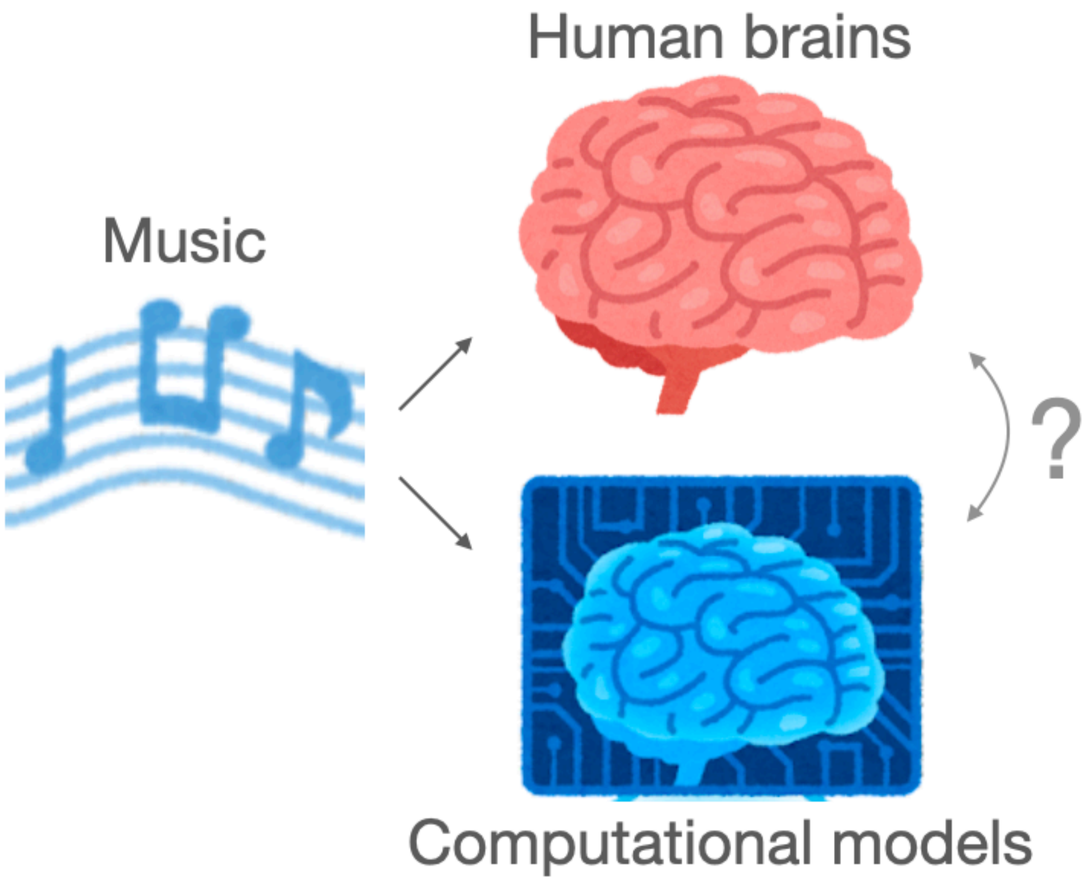
Example: Queen. (1975). Bohemian Rhapsody

VGGish representations at various layers (from 1 to 23)



Audio source: (c) EMI

Research Questions



- **Q1:** Would **increasingly abstract representations of music** in different layers of the CNN be encoded along the cortical gradient axis of abstraction?
- **Q2**: How do **layer-specific CNN embeddings** predict human behavioral ratings of musical emotions?

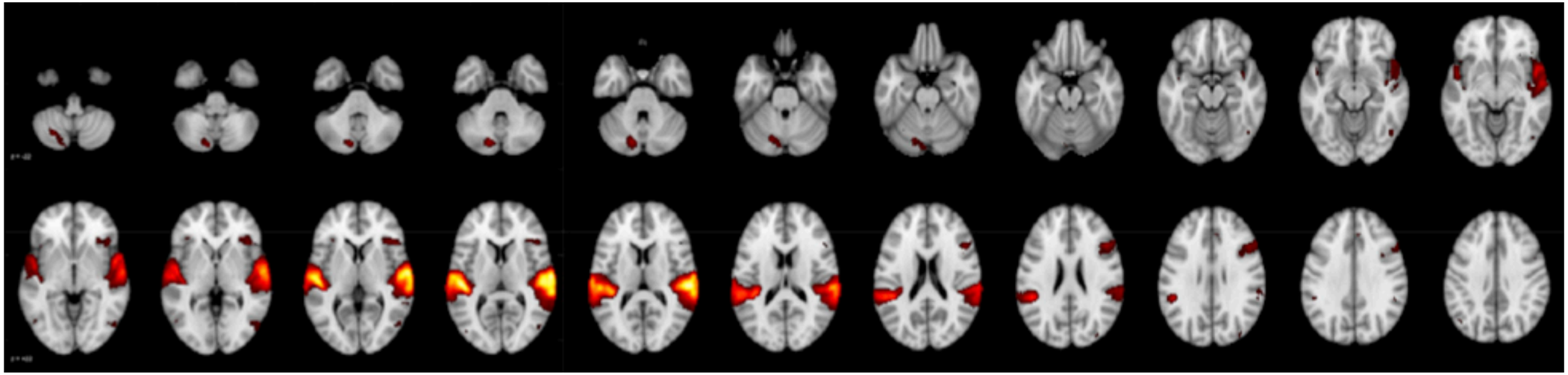
Kim, 2022. *Front. Neurosci.*
https://www.irasutoya.com/

Original study

Sachs et al., 2020, *NeuroImage*.

https://openneuro.org/datasets/ds003085

Inter-subject correlation during a "sad" piece of music: r ~ [0, 0.16], cluster-*P* < 0.05

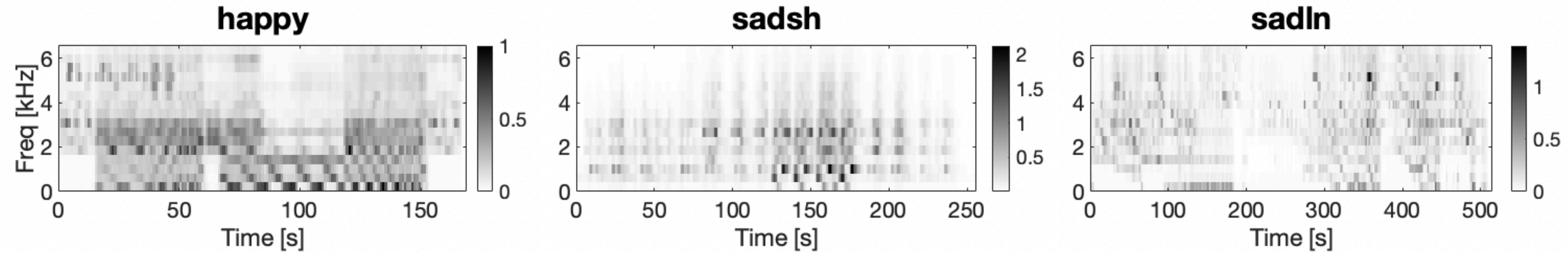


Sachs et al., 2020, *NeuroImage.*

Stimuli

Sachs et al., 2020, *NeuroImage*.

- **Happy** [2 min 48 sec]: Lullatone's "Race against the Sunset"
- **Sad-short** [4 min 16 sec]: Olafur Arnalds's "Frysta"
- **Sad-long** [8 min 35 sec]: Michael Kamen's "Discovery of the Camp"



Kim et al., *In prep*.

Participants & protocol

Sachs et al., 2020, *NeuroImage*.

- N = 40 (21 female, mean age = 24.1 ± 6.24 from LA)
  - Unfamiliar with 3 stimuli and reported "intended" emotions from 60-s excerpts

Passive listening with eyes open



<Not kids: just Japanese illustration of adults>

Rating with a slider



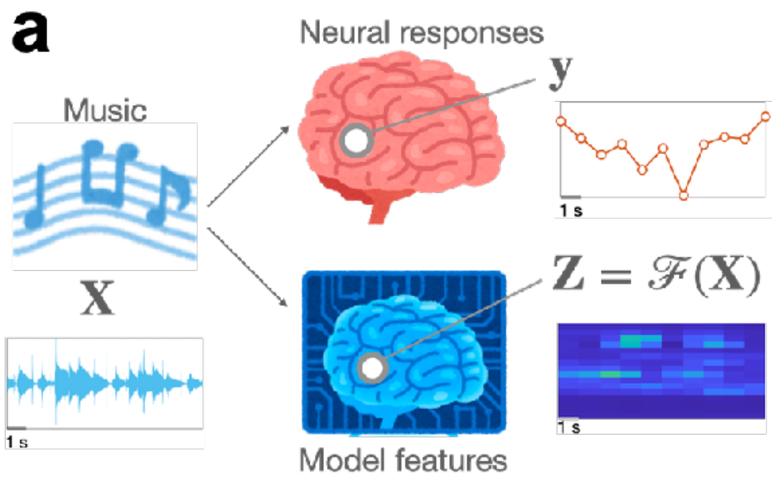
<Not kids: just Japanese illustration of adults>

- "The intensity of felt sadness or happiness" (***Emotionality***)
- "The intensity of enjoyment" (***Enjoyment***)

https://www.irasutoya.com/

Layer-specific encoding analysis

How do we map VGGish-layer-specific encoding on the cortex?

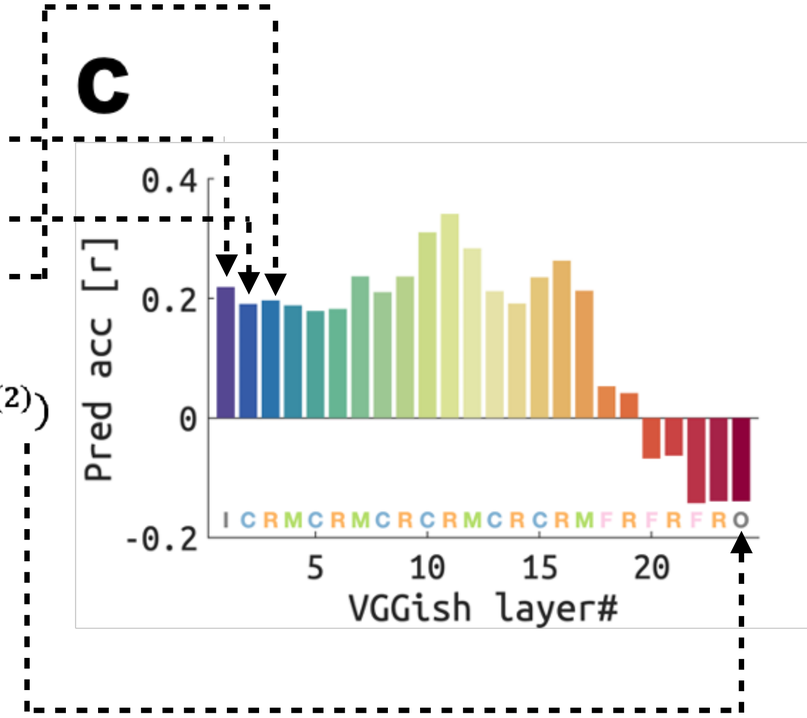


Independent voxel-wise layer-specific encoding models

b

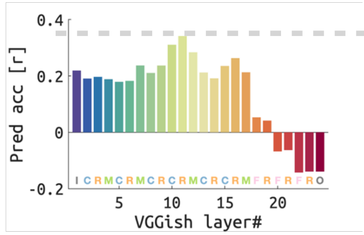
$$\boldsymbol{y}^{(1)} = \boldsymbol{\mathcal{F}}_1(\mathbf{X}^{(1)})\mathbf{b}_1 + \varepsilon \rightarrow \boldsymbol{r}_1 = \mathrm{corr}(\hat{\boldsymbol{y}}_1^{(2)}, \boldsymbol{y}^{(2)})$$
$$\boldsymbol{y}^{(1)} = \boldsymbol{\mathcal{F}}_2(\mathbf{X}^{(1)})\mathbf{b}_2 + \varepsilon \rightarrow \boldsymbol{r}_2 = \mathrm{corr}(\hat{\boldsymbol{y}}_2^{(2)}, \boldsymbol{y}^{(2)})$$
$$\boldsymbol{y}^{(1)} = \boldsymbol{\mathcal{F}}_3(\mathbf{X}^{(1)})\mathbf{b}_3 + \varepsilon \rightarrow \boldsymbol{r}_3 = \mathrm{corr}(\hat{\boldsymbol{y}}_3^{(2)}, \boldsymbol{y}^{(2)})$$
$$\vdots$$
$$\boldsymbol{y}^{(1)} = \boldsymbol{\mathcal{F}}_{24}(\mathbf{X}^{(1)})\mathbf{b}_{24} + \varepsilon \rightarrow \boldsymbol{r}_{24} = \mathrm{corr}(\hat{\boldsymbol{y}}_{24}^{(2)}, \boldsymbol{y}^{(2)})$$

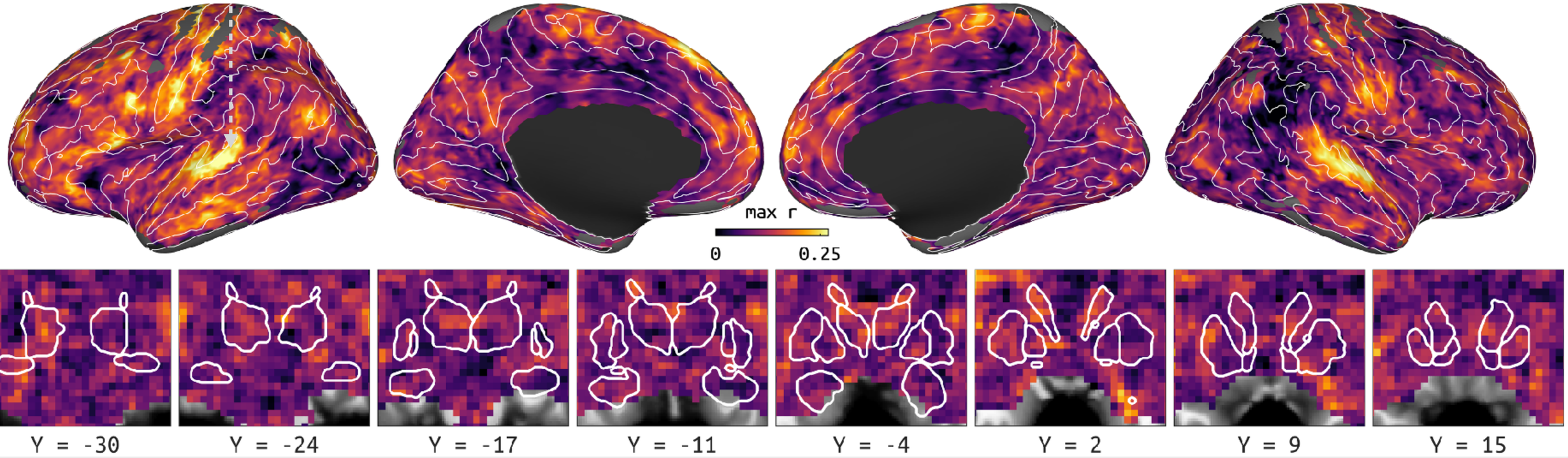
Prediction accuracies across layers of a voxel

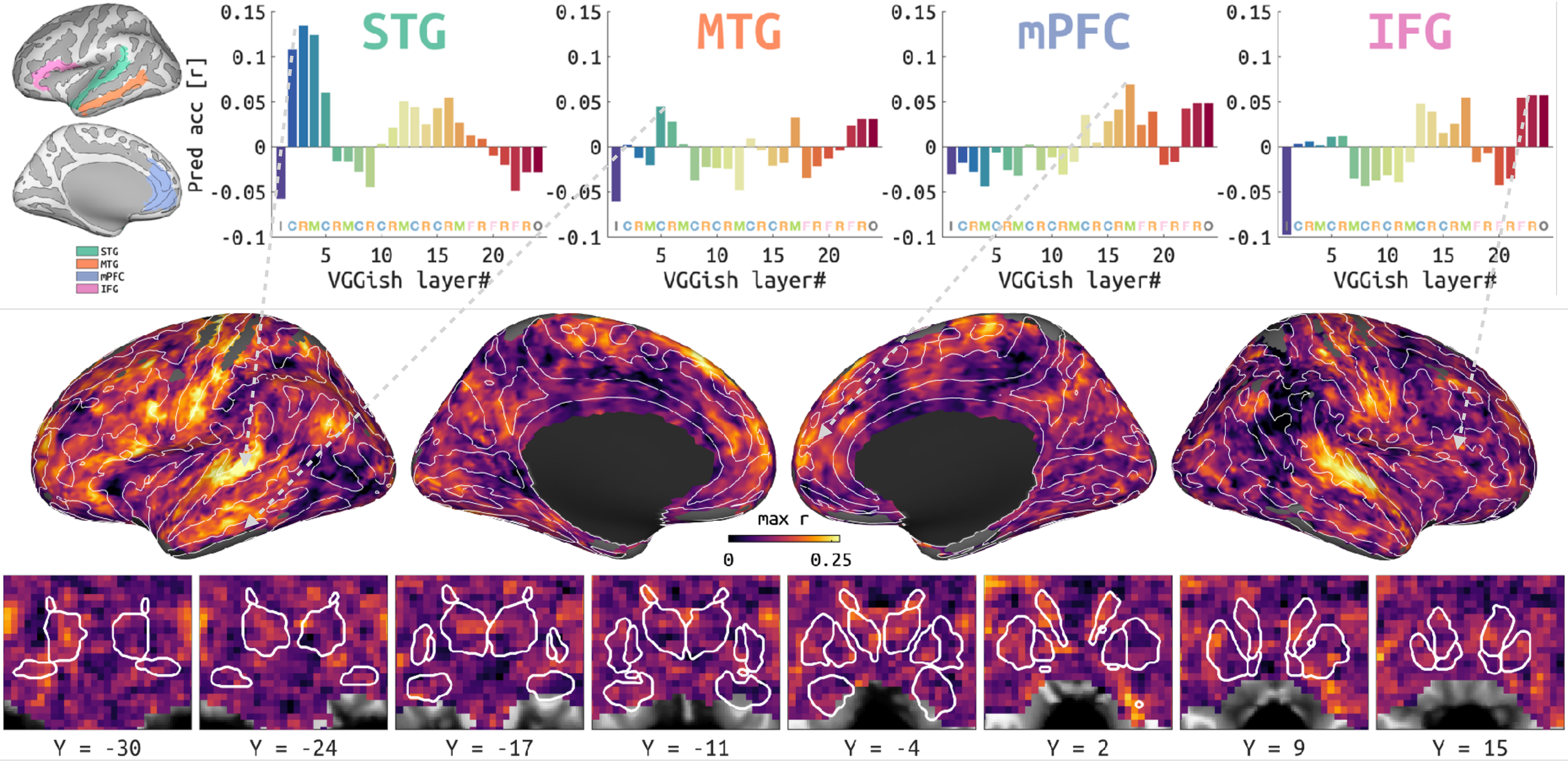


Maximal prediction accuracy 🎶→🧠

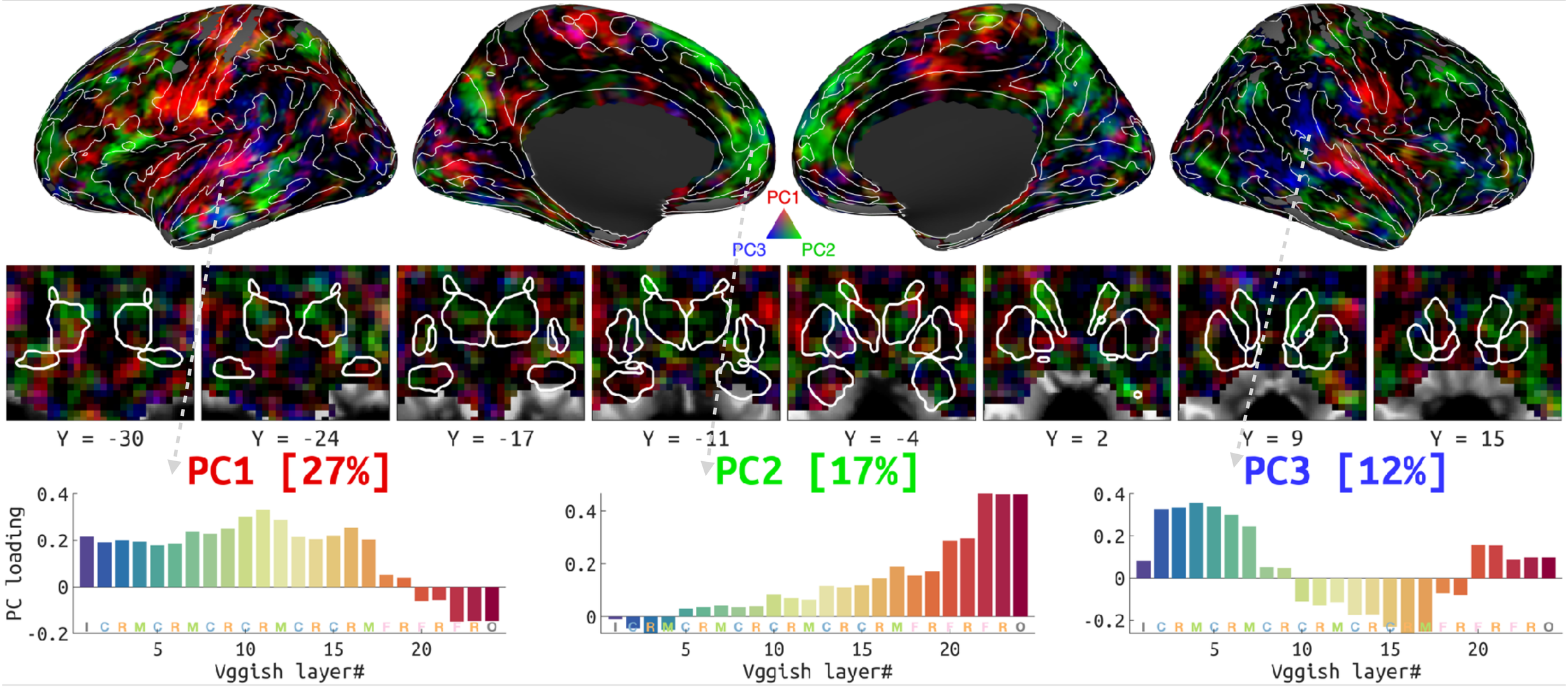
c



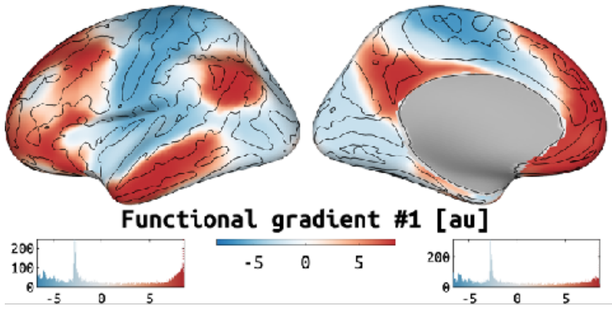



STG
MTG
mPFC
IFG
Pred acc [r]
0.15
0.1
0.05
0
-0.05
-0.1
I C R M C R M C R C R M C R C R M F R F R F R O
5
10
15
20
VGGish layer#
max r
0
0.25
Y = -30
Y = -24
Y = -17
Y = -11
Y = -4
Y = 2
Y = 9
Y = 15

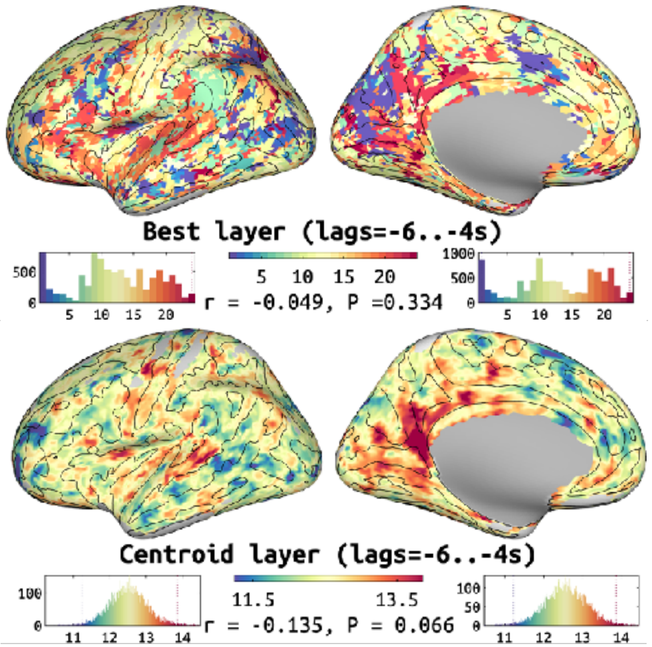
Topography of CNN-layer-specific encoding 🎶→🧠

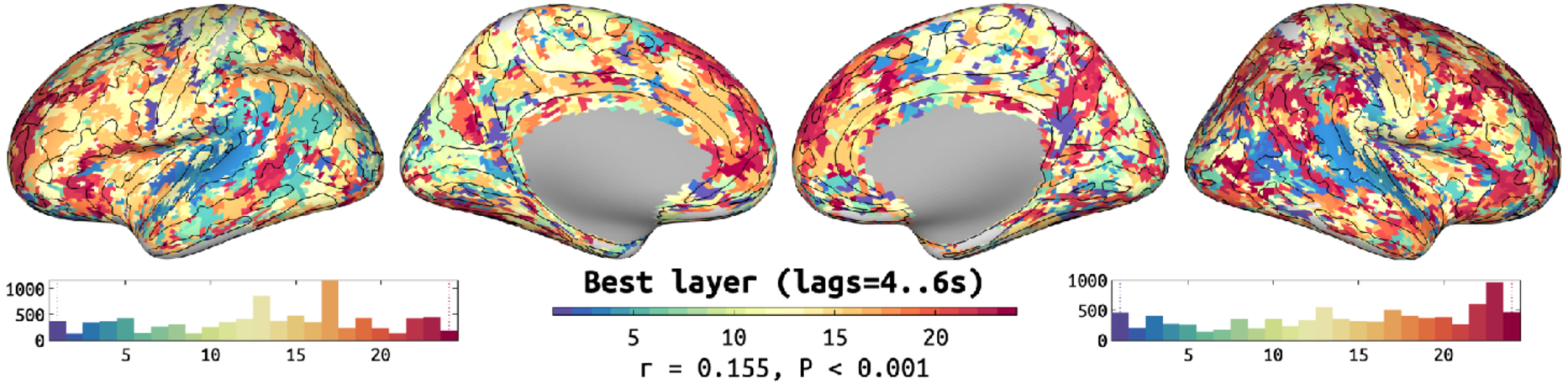


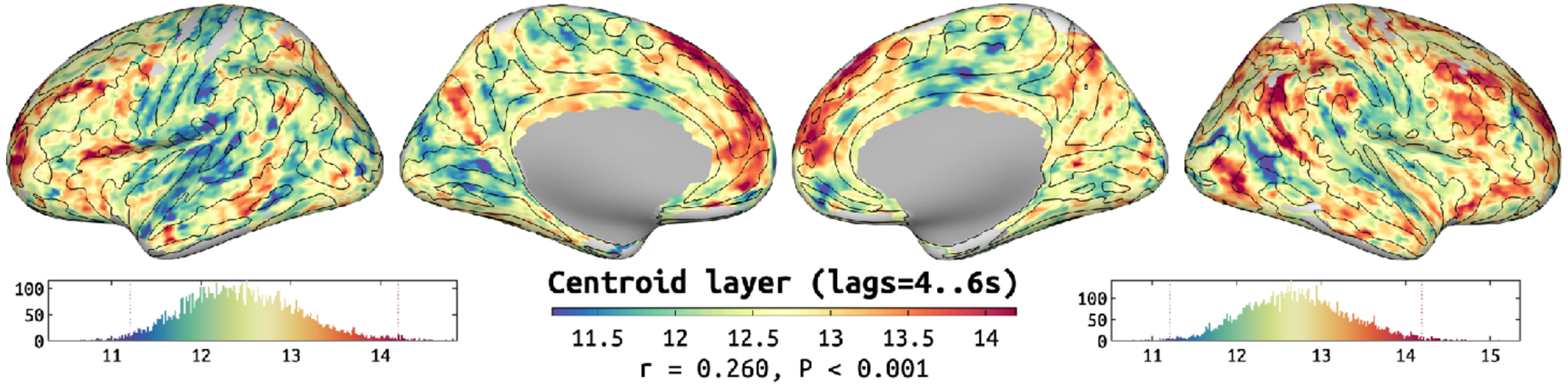
RSFC G1 vs. the "best" CNN-layer map



Is it "*just*" because of fMRI noise structures? 🧐



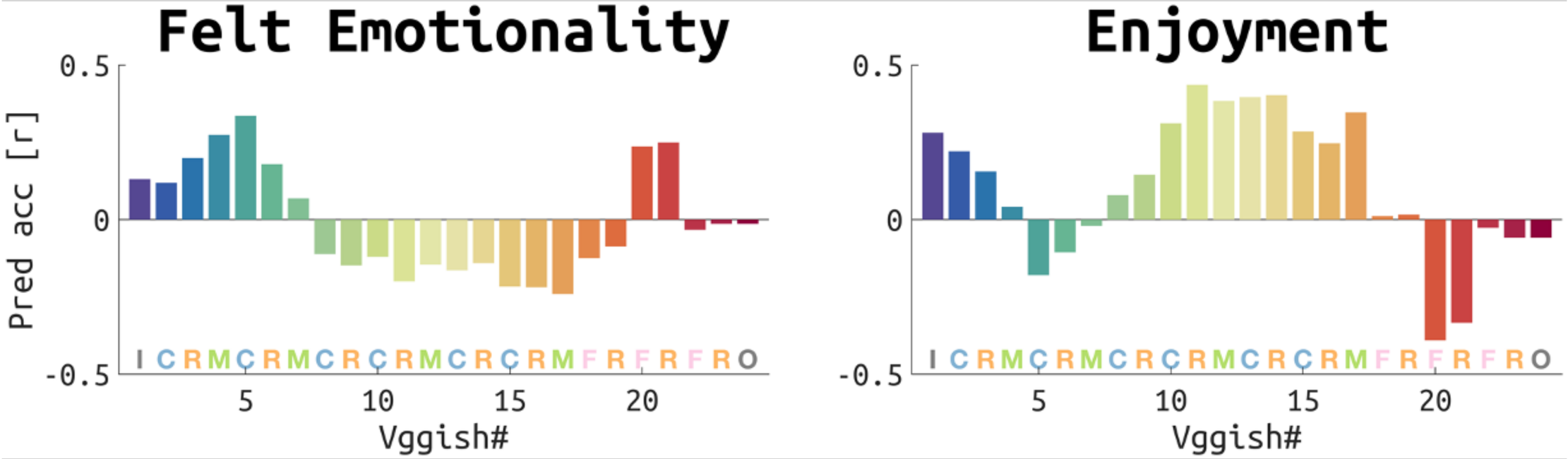




Deeper CNN-layer embedding is encoded in "transmodal" cortical areas.

Distinctive patterns of emotional ratings

🎶→😀😢



Any questions so far?

Workshop plan

Building a 1,000-Song Naturalistic EEG Dataset

Part	From	To	Contents	Dur
1	09:00	09:45	**Introduction**	45m
2	10:00	10:45	**Analysis**	45m
3	11:00	11:45	**Design**	45m
4	12:00	12:45	**Management**	45m

Methods (WHYs):
Linearised Encoding Analysis

How can we understand a nonlinear system?

Let's say we magically know what the brain does:

$$\underset{(t\times\psi)}{\mathbf{S}} \qquad \mathcal{F} : \mathbb{R}^{\psi} \mapsto \mathbb{R}^{v} \qquad \underset{(t\times v)}{\mathbf{Y}}$$



Stimulus



System

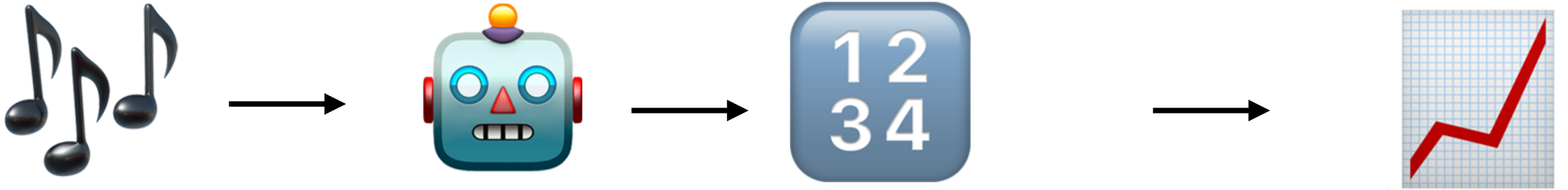


Response

$$\underset{(t\times v)}{\mathbf{Y}} = \underset{(t\times v)}{\mathcal{F}(\mathbf{S})} + \underset{(t\times v)}{\mathbf{E}}$$

Let's also say we (somehow) know a nonlinear function that allows us a linear mapping

$$\underset{(t\times\psi)}{\mathbf{S}} \quad \mathcal{G} : \mathbb{R}^{\psi} \mapsto \mathbb{R}^{f} \quad \underset{(t\times f)}{\mathbf{X}} \quad \mathcal{B} : \mathbb{R}^{f} \mapsto \mathbb{R}^{v} \quad \underset{(t\times v)}{\mathbf{Y}}$$



Stimulus → Model → Features → Response

$$\underset{(t\times v)}{\mathbf{Y}} = \underset{(t\times f)}{\mathcal{G}(\mathbf{S})} \; \underset{(f\times v)}{\mathbf{B}} + \underset{(t\times v)}{\mathbf{E}}$$

Linearised encoding analysis

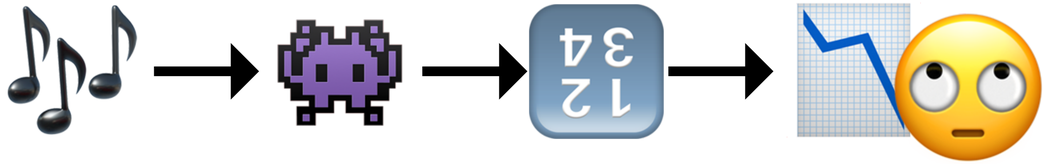
A linear time-invariant system identification

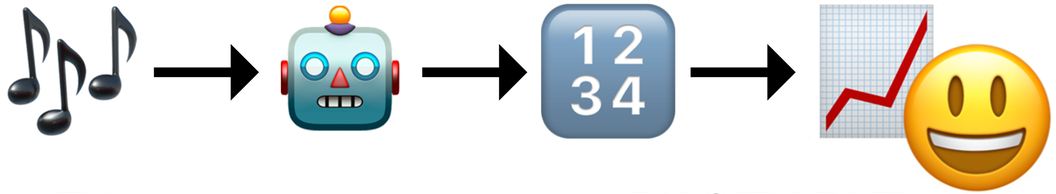
- The method is widely known as "system identification": we separate a nonlinear transform (stim→resp) into a simpler nonlinear transform (stim→feat) and a linear transform (feat→resp).
- How the stimulus is *linearised* (by which function) is <u>your hypothesis.</u>
- Then, the rest is to recover a linear filter ("convolution kernel") that transforms features $\mathcal{G}(\mathbf{S})$ to response $\mathbf{Y}$.

$$\mathbf{Y} = \mathcal{F}(\mathbf{S}) + \varepsilon$$

$$\mathbf{Y} = \mathcal{G}(\mathbf{S})\mathbf{B} + \varepsilon$$







This means the hypothesis is ***FALSIFIABLE!***

👾: ❌ but 🤖: ✅

🤔: But what if the internal process takes unknown time?

Topic 1: Finite impulse response modeling

Modeling (neural) system delays

$$y_t = b_0 + x_t b_1 + \epsilon$$

y(t) is only influenced by x(t)

Topic 1: Finite impulse response modeling

Modeling (neural) system delays

$$y_t = b_0 + x_t b_1 + \epsilon$$

y(t) is only influenced by x(t)

$$y_t = b_0 + x_t b_1 + x_{t-1} b_2 + \epsilon$$

y(t) is influenced by x(t) and x(t-1)

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_0 \\ 1 & x_2 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_{n-1} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Topic 1: Finite impulse response modeling

Modeling (neural) system delays

$$y_t = b_0 + x_t b_1 + \epsilon$$

y(t) is only influenced by x(t)

$$y_t = b_0 + x_t b_1 + x_{t-1} b_2 + \epsilon$$

y(t) is influenced by x(t) and x(t-1)

$$y_t = b_0 + x_t b_1 + x_{t-1} b_2 + x_{t-2} b_3 + \epsilon$$

y(t) is influenced by x(t:t-2)

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_0 & x_{-1} \\ 1 & x_2 & x_1 & x_0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_{n-1} & x_{n-2} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Topic 1: Finite impulse response modeling

Modeling (neural) system delays

$$y_t = b_0 + x_t b_1 + \epsilon$$ **y(t) is only influenced by x(t)**

$$y_t = b_0 + x_t b_1 + x_{t-1} b_2 + \epsilon$$ **y(t) is influenced by x(t) and x(t-1)**

$$y_t = b_0 + x_t b_1 + x_{t-1} b_2 + x_{t-2} b_3 + \epsilon$$ **y(t) is influenced by x(t:t-2)**

$$\vdots$$

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \boldsymbol{\varepsilon}$$

<#Timepoints x [#delays+1]>

<#Timepoints x 1>

<#Timepoints x 1>

<[#delays+1] x 1>

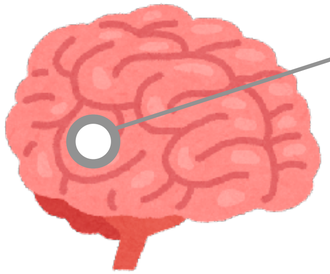
- This is a discrete equivalent of convolution (thus, the vector **b** is also called a "filter" or a "kernel").
- "**Finite**": Because we model the impulse response up to #delay.
- "**Impulse response**": A response to an impulse of a dynamic system

🤔: But does it work for the other samples?

Topic 2 & 3: Predictive modelling

Training set

Neural responses



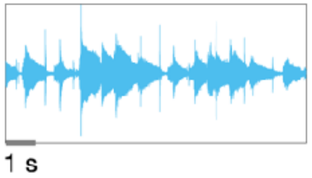
$\mathbf{y}_1$



Music #1



$\mathbf{S}_1$

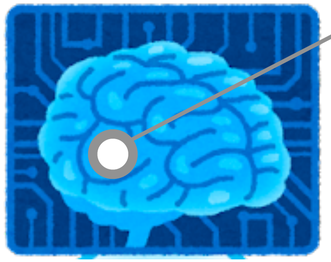


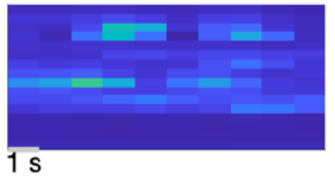
Weight fitting: $\hat{\mathbf{b}}_1 = (\mathbf{X}_1'\mathbf{X}_1 + \mathbf{L}^*)^{-1}\mathbf{X}_1'\mathbf{y}_1$

$\mathbf{y}_1 = \mathbf{X}_1\mathbf{b}_1 + \varepsilon$

FIR delays

$\mathbf{x}_1 = \mathcal{G}(\mathbf{S}_1)$

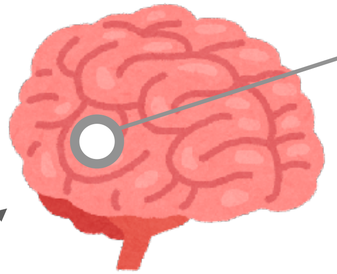




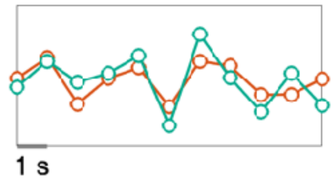
Model features

Test set

Neural responses



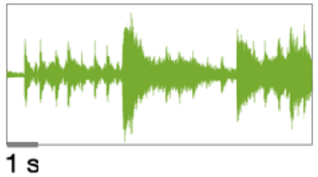
$\mathbf{y}_2$



Music #2



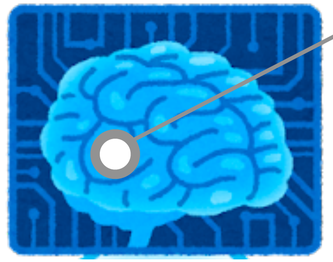
$\mathbf{X}_2$



Prediction: $\hat{\mathbf{y}}_2 = \mathbf{X}_2\hat{\mathbf{b}}_1$

FIR Delays

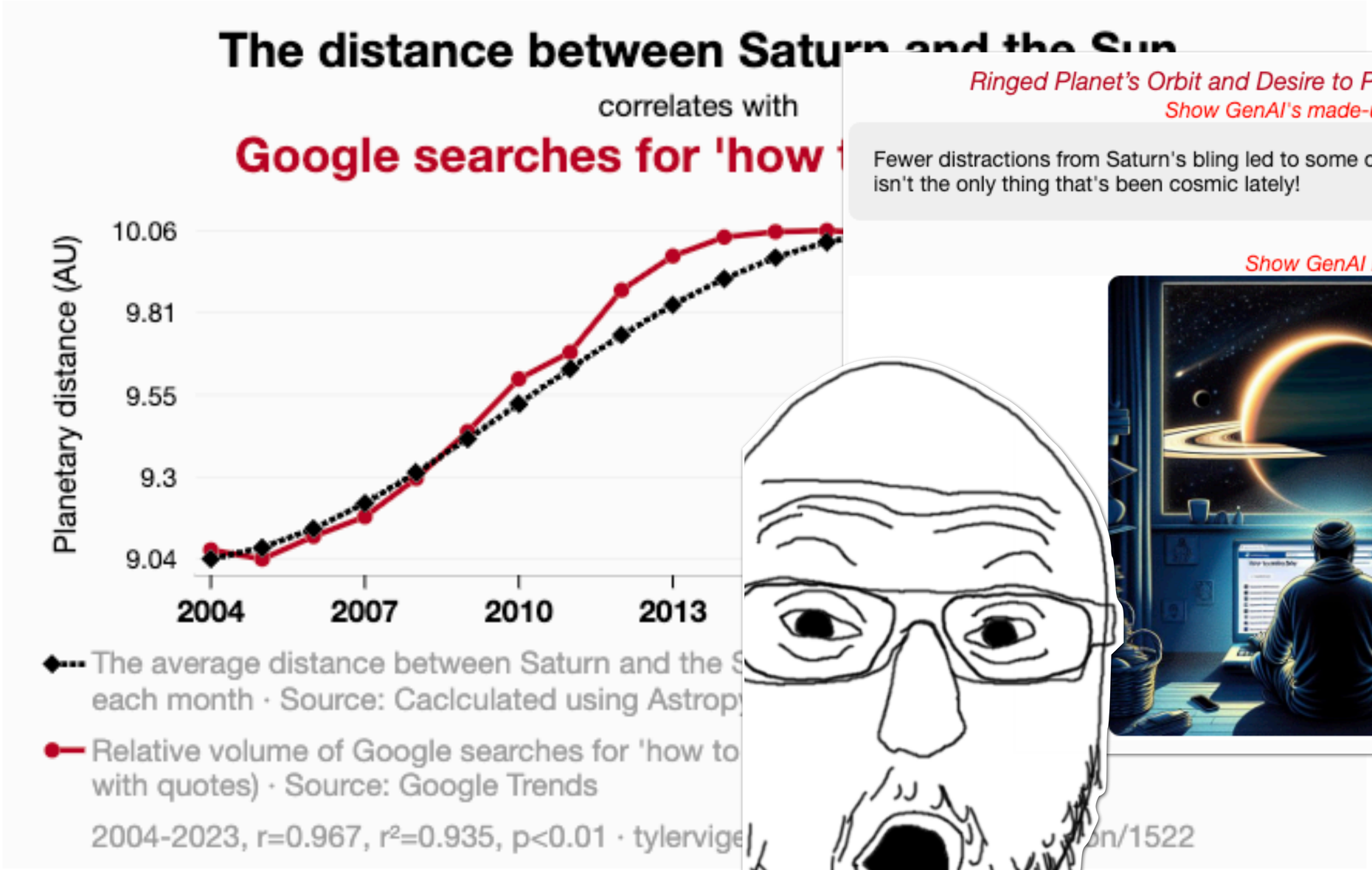
$\mathbf{x}_2 = \mathcal{G}(\mathbf{S}_2)$





Model features

🤔: Still, how do we know if this is not due to the random noise in the data?



Ringed Planet's Orbit and Desire to Procreate: A Statistical Analysis

Show GenAI's made-up explanation

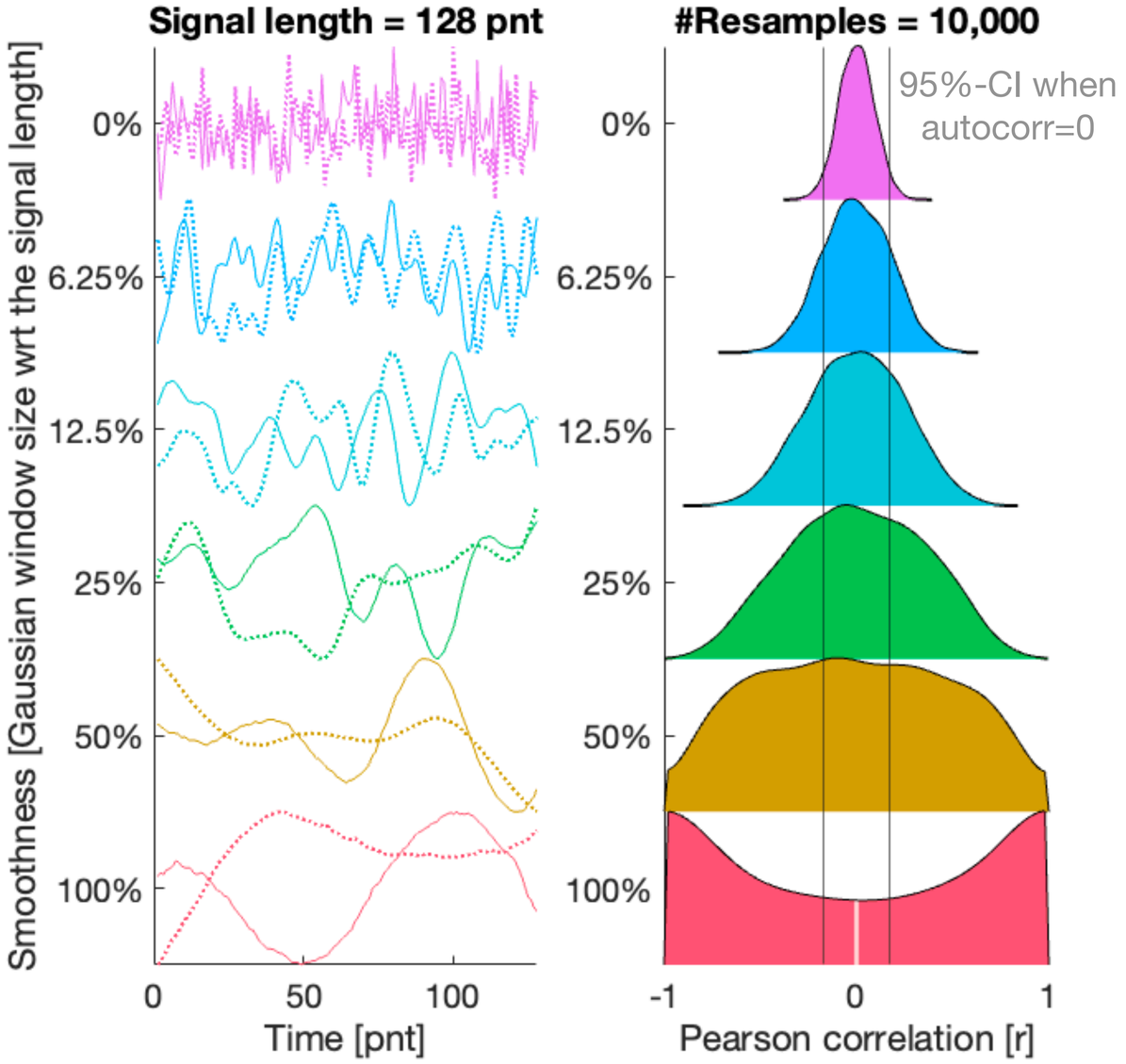
Fewer distractions from Saturn's bling led to some out-of-this-world romance. Looks like love isn't the only thing that's been cosmic lately!

Show GenAI image

Topic 4: Statistical inference

Tests for time series

- Many **time series** look alike (i.e., spurious correlation) due to auto/ serial-correlation.
- **Non-parametric statistical tests are feasible for any arbitrary noise distribution** with "proper" null (surrogate) data.
- Explicitly estimation autocorrelation from finite (non-infinite) data is hard.
- Preserving the dependency structure (e.g., serial/spatial, ...) is rather easy.



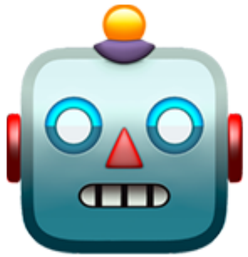
Summary: linearised encoding analysis



Stimulus

$\mathcal{G}(\mathbf{S})$

Feature extraction models
- Music information retrieval
- Computational musicology
- Auditory/music psychology
- ...



Model



Human



Null model



Features



Null feat.

FIR

Reg



Response



Prediction



Null pred.

CV: opt & eval

Inference

Any questions so far?

Workshop plan

Building a 1,000-Song Naturalistic EEG Dataset

Part	From	To	Contents	Dur
1	09:00	09:45	**Introduction**	45m
2	10:00	10:45	**Analysis**	45m
3	11:00	11:45	**Design**	45m
4	12:00	12:45	**Management**	45m

ManyMusic🎶
motivation & design

Limitations of previous approaches

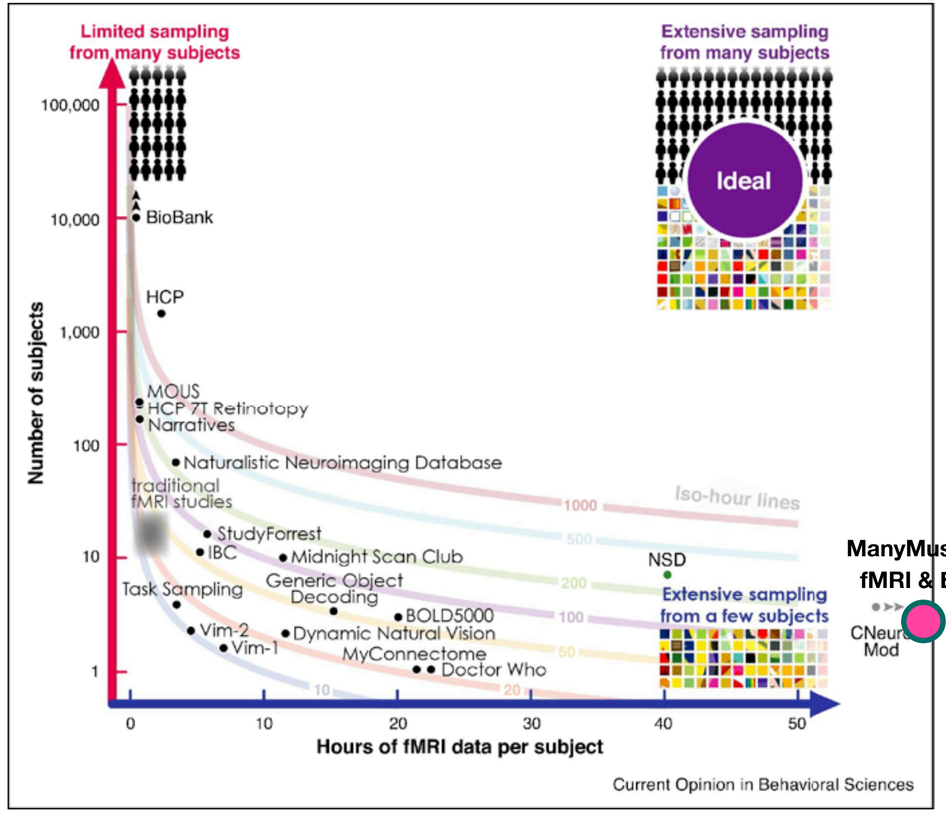
- **Traditional neuroscientific approaches…**
  - used a small number of short musical excerpts (mainly classical),
  - often compared with artificially manipulated stimuli (not without artefacts),
  - relied on group-statistical analyses of averaged brain activity.

- **Difficulties in applying to real-world music?**
  - *Enormous dynamics and complexity* of emotional experiences during natural music listening [1]
  - *Variability* of the music people hear, know, and like [2]
  - *Uniqueness* of individual neural functioning [3]

[1] Cowen et al., 2020, *PNAS* [2] Park et al., 2019, *Nat Hum Beh* [3] Westlin et al., 2023, *TiCS*

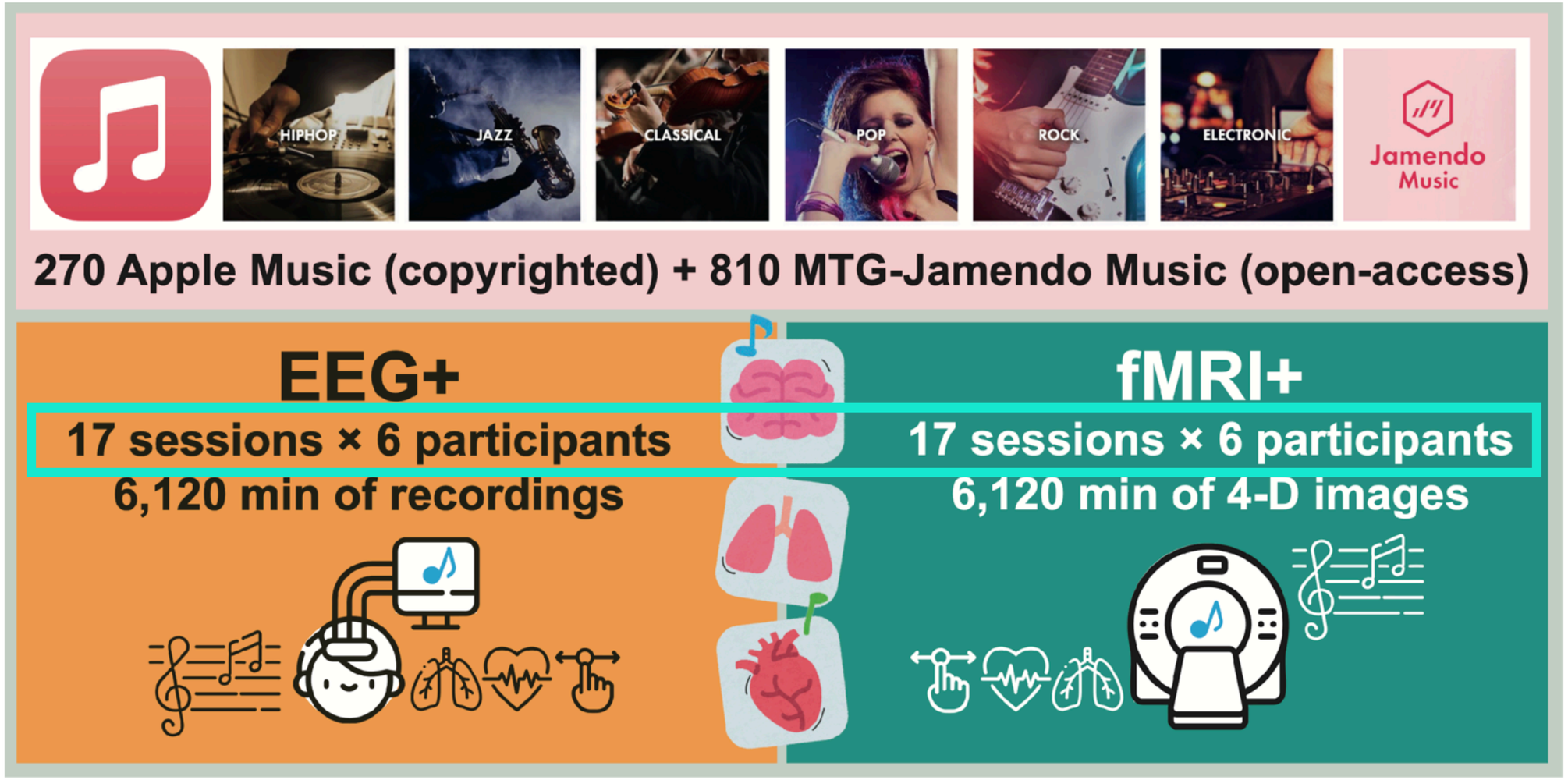
Aims of ManyMusic🎶

- **Stimulus**: ***Full-length*** presentation of music in ***diverse genres*** eliciting ***various emotions***
- **Measurement**: ***Extensive sampling*** of subjective experience of music-evoked affection in a selected individuals to minimize individual variability in the dataset
- **Accessibility**: ***Fully open-access*** dataset of stimuli, computational features, and behavioural/ neurophysiological responses
- **Expected outcome**: ***Precise predictive modelling of musical emotions***



Naselaris et al., 2020, *Cur Opin in Beh Sci*

Design and Components



https://flaticon.com

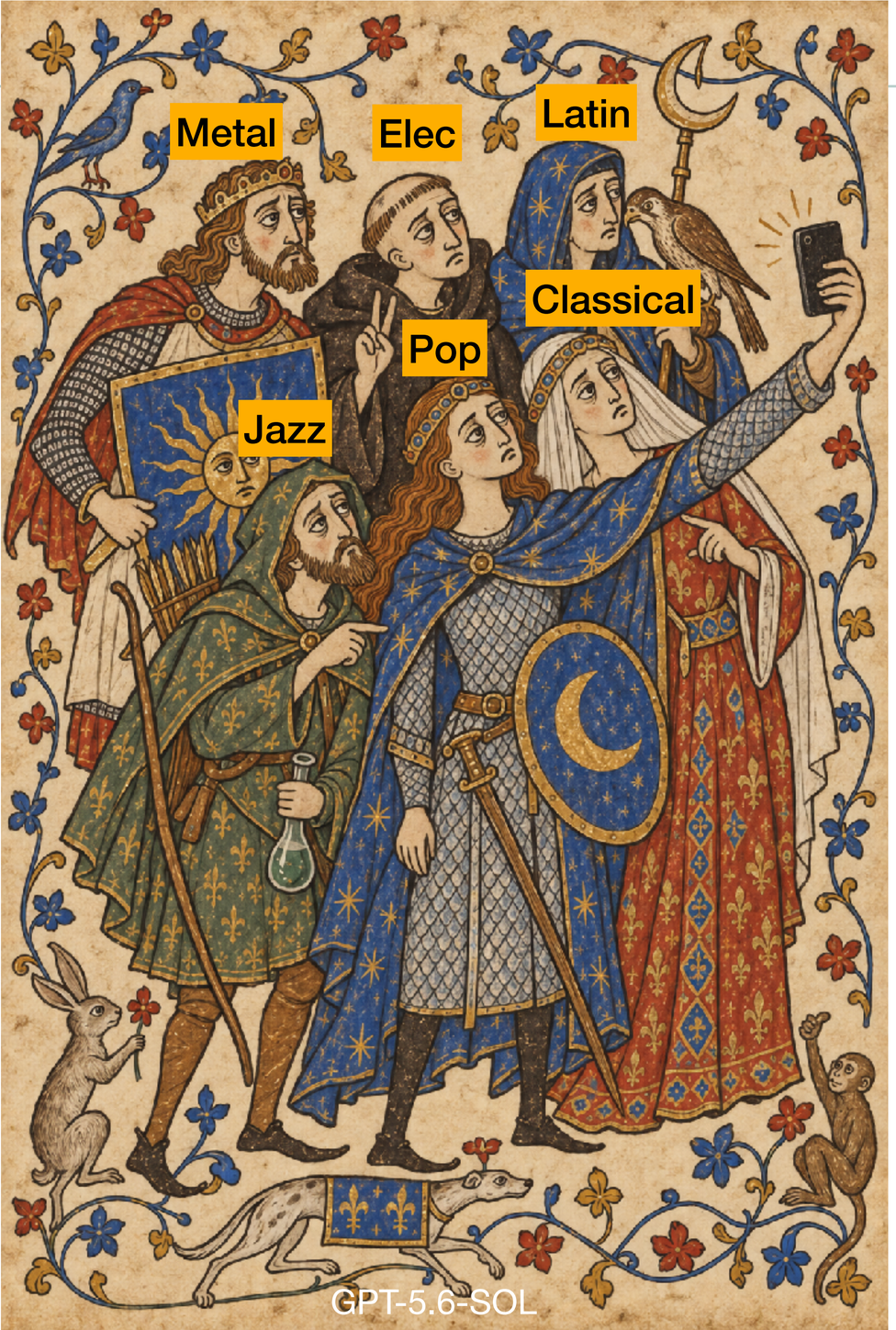
Participant screening

Multiple steps to find the "*Hero Participants*"

- **We need six music lovers who can do 34+ sessions.**
- **Online Screening Survey**: N = 48 (3 weeks)
  - Musical sophistication (G-MSI), musical reward (BMRQ), personality (10-item Big-5), …
- **Online Structured Interview**: N = 24 (2 weeks)
  - Interest in music, high commitment, …
- **fMRI Screening Sessions**: N = 12 (2 weeks)
  - Test-retest reliability, high sensitivity to music, least head motion
- **Final Selection**: N = 6
  - A couple of musicians dropped out because of the scanner noise (even after ANC) and the loud (>80 SPL) stimulus presentation.



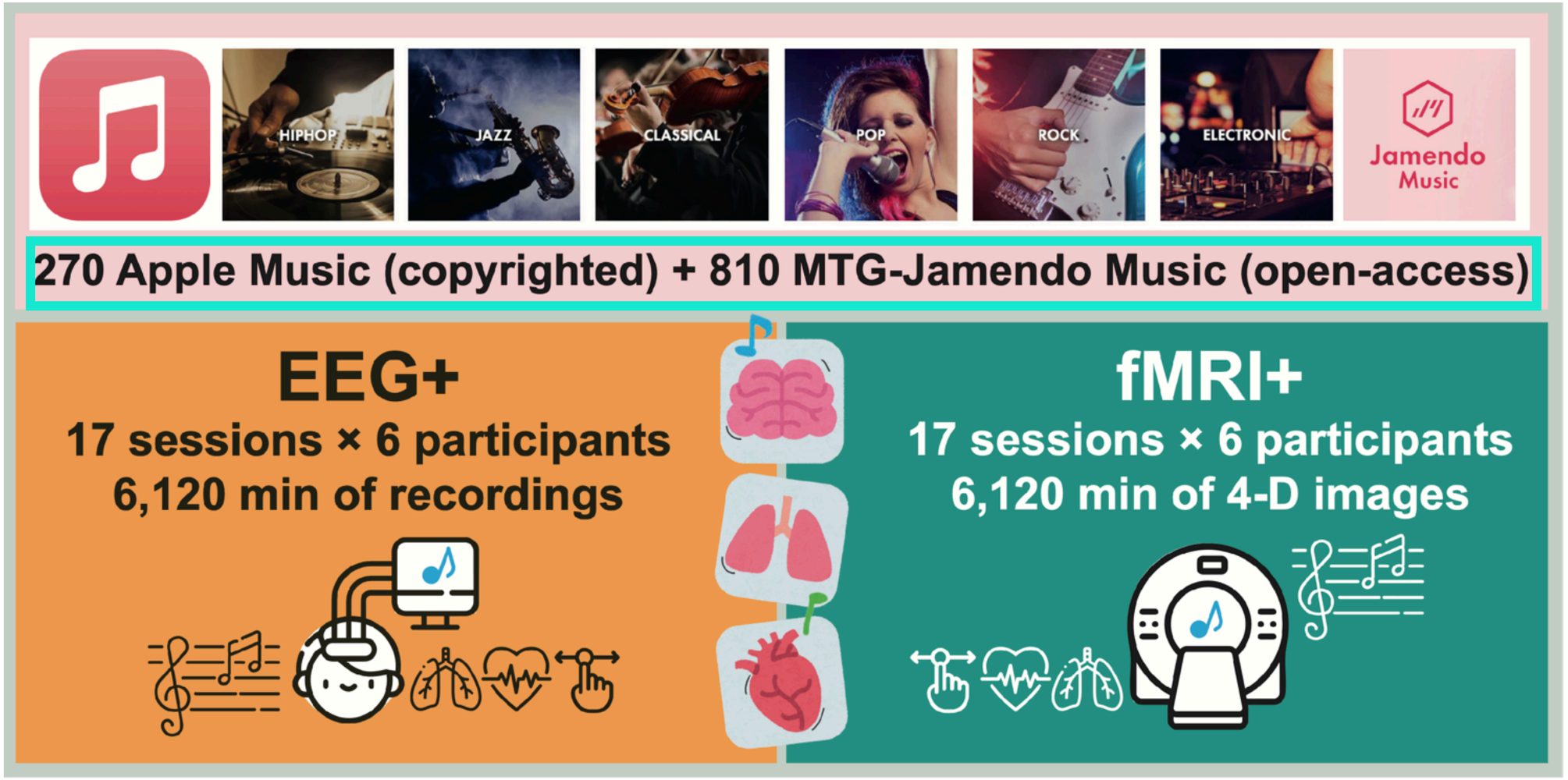
GPT-5.6-SOL



Six legendary champions

- Multicultural (German, Columbian, Indian, Romanian) and multi-genre (classical, jazz, metal, rock, electronica, pop)
- **0% dropout after 11 months**
- Exponentially increasing return bonus (4 to 100€) plus base rates
- Good relationship with participants
- Flexible scheduling (e.g., 2-3w vacation)
- Good selection and luck!

Design and Components



https://flaticon.com

Representative music datasets

Table 1

Representative previous datasets.

Discipline	Dataset name	#Tracks	Duration	#Genres	Source	Accessibility
Psychology	Film Soundtrack [12]	360	10–60 sec	1	Commercial	Unclear
Psychology	EMMA [6]	817	2-7 min	7	Commercial	Copyrighted
Psychology	MUSIFEAST-17 [16]	356	30 sec	17	Commercial	Open-access
MIR	GTZAN [17]	1,000	30 sec	10	Commercial	Unclear
MIR	RWC [13]	215	3-7 min	10	Commissioned	On-request
MIR	FMA [14]	106,000	0-30 min	161	Community	Open-access
MIR	MTG-Jamendo [15]	55,094	0-30 min	95	Community	Open-access

[6] H. Strauss, J. Vigl, P.-O. Jacobsen, M. Bayer, F. Talamini, W. Vigl, E. Zangerle, M. Zentner, The emotion-to-music mapping atlas (EMMA): A systematically organized online database of emotionally evocative music excerpts, Behavior Research Methods 56 (2024) 3560–3577.

[12] T. Eerola, J. K. Vuoskoski, A comparison of the discrete and dimensional models of emotion in music, Psychology of Music 39 (2011) 18–49.

[13] M. Goto, H. Hashiguchi, T. Nishimura, R. Oka, RWC music database: Popular, classical and jazz music databases., in: ISMIR, volume 2, 2002, pp. 287–288.

[14] M. Defferrard, K. Benzi, P. Vandergheynst, X. Bresson, FMA: A dataset for music analysis, in: ISMIR, 2017. arXiv:1612.01840.

[15] D. Bogdanov, M. Won, P. Tovstogan, A. Porter, X. Serra, The MTG-Jamendo dataset for automatic music tagging, in: Machine learning for music discovery workshop, international conference on machine learning (ICML 2019), 2019, pp. 1–3.

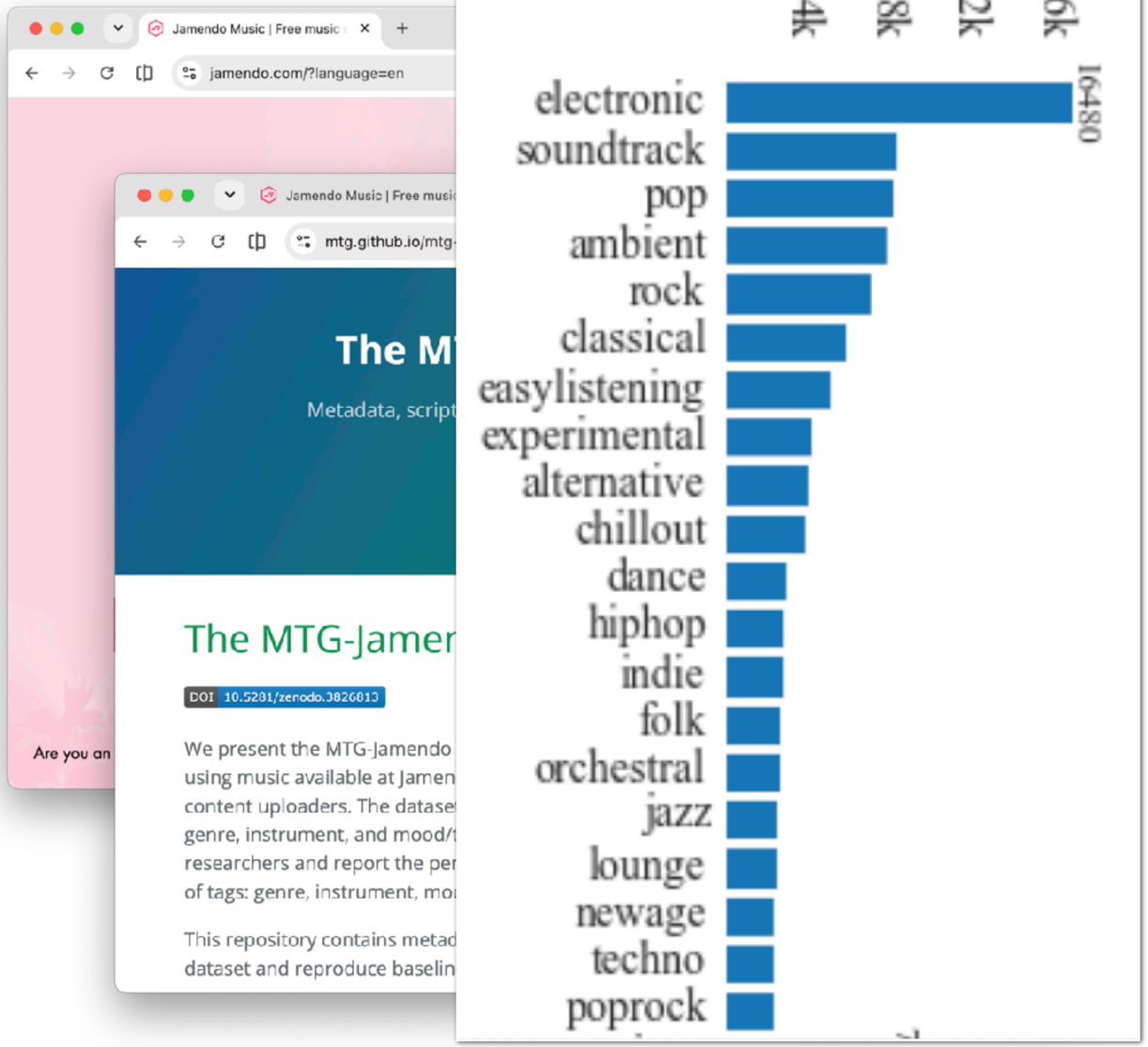
[16] H. A. van der Walle, W. Wu, E. H. Margulis, K. Jakubowski, MUSIFEAST-17: MUsic Stimuli for Imagination, Familiarity, Emotion, and Aesthetic STudies across 17 genres, Behavior Research Methods 57 (2025) 204.

[17] G. Tzanetakis, P. Cook, Musical genre classification of audio signals, IEEE Transactions on Speech and Audio Processing 10 (2002) 293–302. doi:10.1109/TSA.2002.800560.

Open-access music refinement

MTG-Jamendo

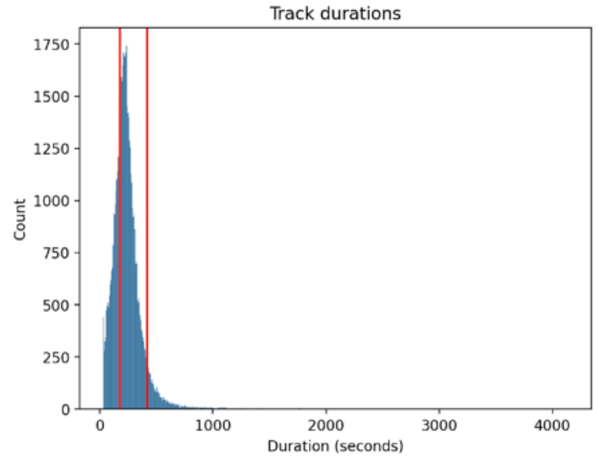
- “Jamendo Music” platform
- Music Technology Group (MTG) published an open-access music dataset (freely available for research)
- 56,639 full-length tracks in 87 genres (but heavily biased towards electronic)



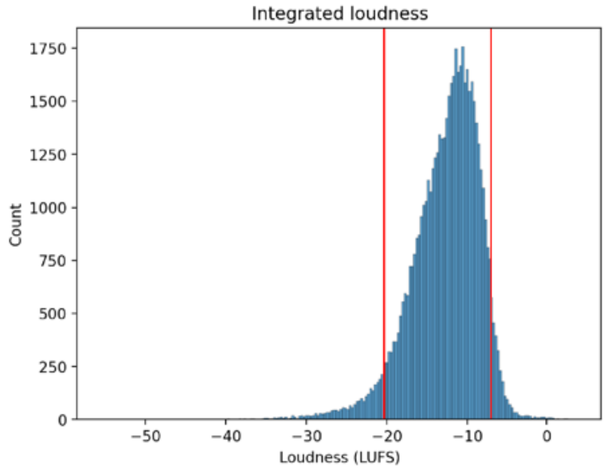
Open-access music refinement

Music filtering

- Duration: 3–7 min
- Loudness: 5–95%
- Clipping artefacts
- Genre Activation ≥ 0.1 for any of 18 genres
- 170 tracks per genre centroid (k=12)

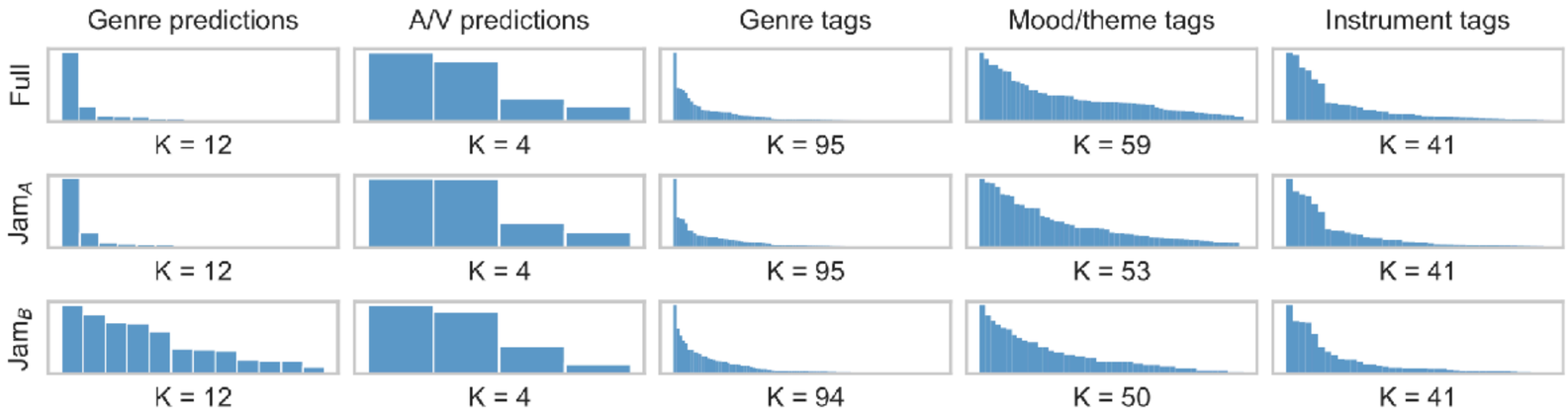


Track duration histogram. The red lines represent the 3 and 7 minute thresholds.

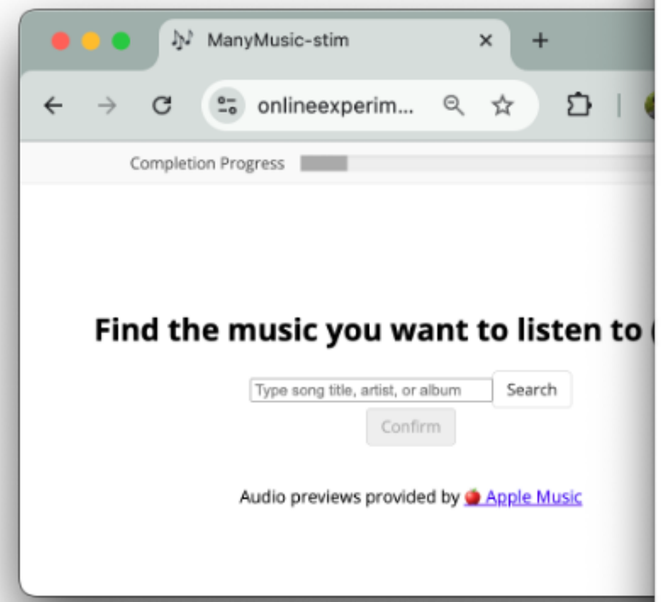


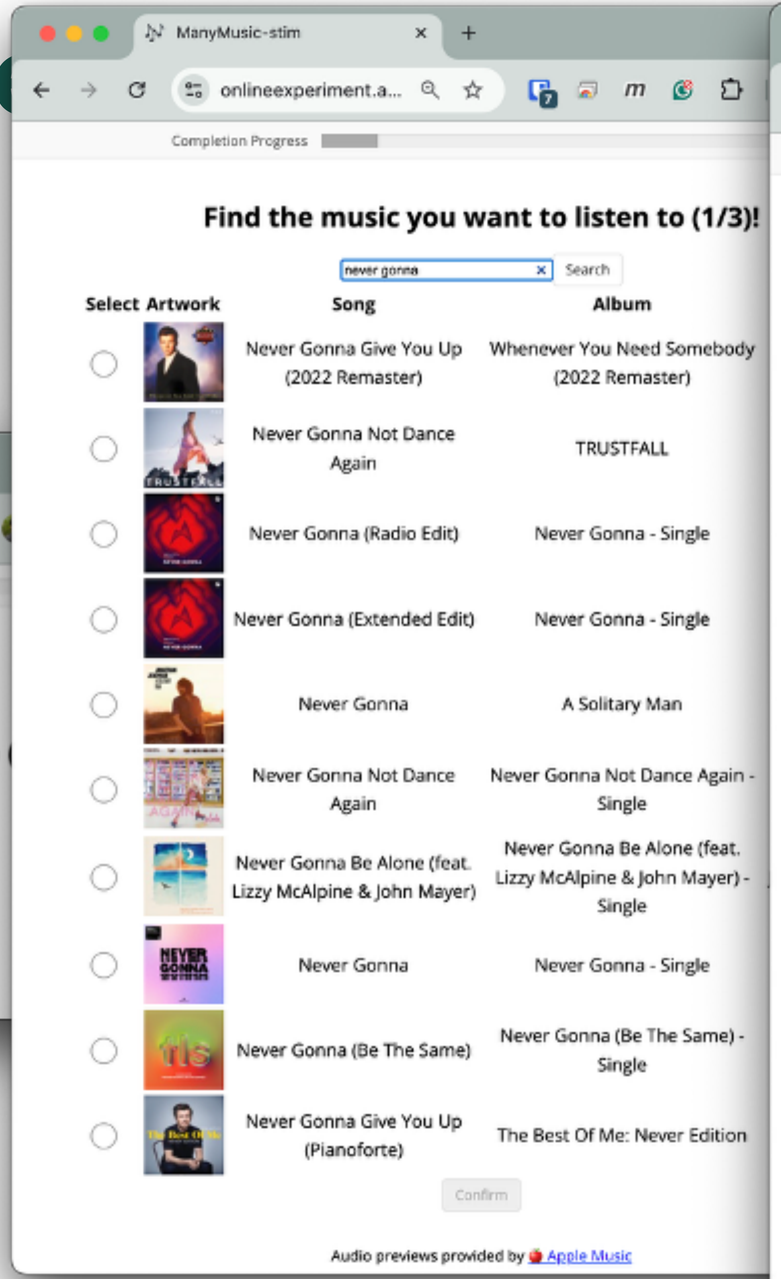
Integrated loudness histogram. The red lines represent the 5th and -95th percentiles of the loudness distribution.

Figure 1: Distribution of model predictions and original tags for the Full dataset and the Jamendo-A and Jamendo-B subsets. Arousal and valence (A/V) predictions are grouped in the four positive/negative A/V quadrants. K denotes the total number of unique classes in each case. See https://manymusic.net/stim/ for more information.



Online expe


ManyMusic-stim
onlineexperim...
Completion Progress
Find the music you want to listen to
Type song title, artist, or album
Search
Confirm
Audio previews provided by Apple Music


ManyMusic-stim
onlineexperiment.a...
Completion Progress
Find the music you want to listen to (1/3)!
never gonna
Search
Select Artwork
Song
Album
Never Gonna Give You Up (2022 Remaster)
Whenever You Need Somebody (2022 Remaster)
Never Gonna Not Dance Again
TRUSTFALL
Never Gonna (Radio Edit)
Never Gonna - Single
Never Gonna (Extended Edit)
Never Gonna - Single
Never Gonna
A Solitary Man
Never Gonna Not Dance Again
Never Gonna Not Dance Again - Single
Never Gonna Be Alone (feat. Lizzy McAlpine & John Mayer)
Never Gonna Be Alone (feat. Lizzy McAlpine & John Mayer) - Single
Never Gonna
Never Gonna - Single
Never Gonna (Be The Same)
Never Gonna (Be The Same) - Single
Never Gonna Give You Up (Pianoforte)
The Best Of Me: Never Edition
Confirm
Audio previews provided by Apple Music


ManyMusic-stim
onlineexperiment....
Completion Progress
Unfamiliar
This music sounds ...
Familiar
Moved, Touched, or Engaged
I feel ...
Bored, Disgusted, or Disengaged
Unprofessional
This music sounds ...
Professional
Like it
I do ...
Dislike it
Joyful, Amusing, or Other Positive Feelings
This music expresses ...
Melancholic, Agitating, or Other Negative Feelings
Exciting, Stimulating, or Energizing
This music is ...
Calming, Soothing, or Relaxing
Please move the sliders above after listening to the music!
Continue
Cancel

Open-access music refinement

Selection validation

- Jamendo-A: N = 2599
- Jamendo-B: N = 1119 (3-raters)
- Spotify previews: N = 2100
- Suno-v4.5: N = 1280
- Self-selected: N = 3 (per subject)

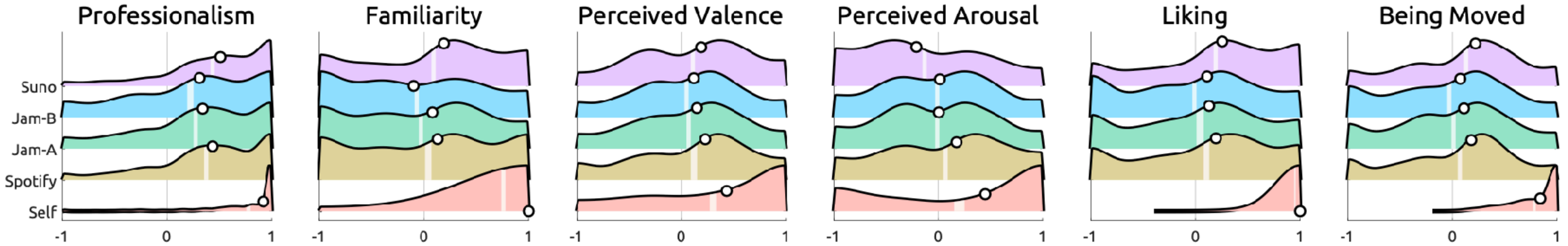
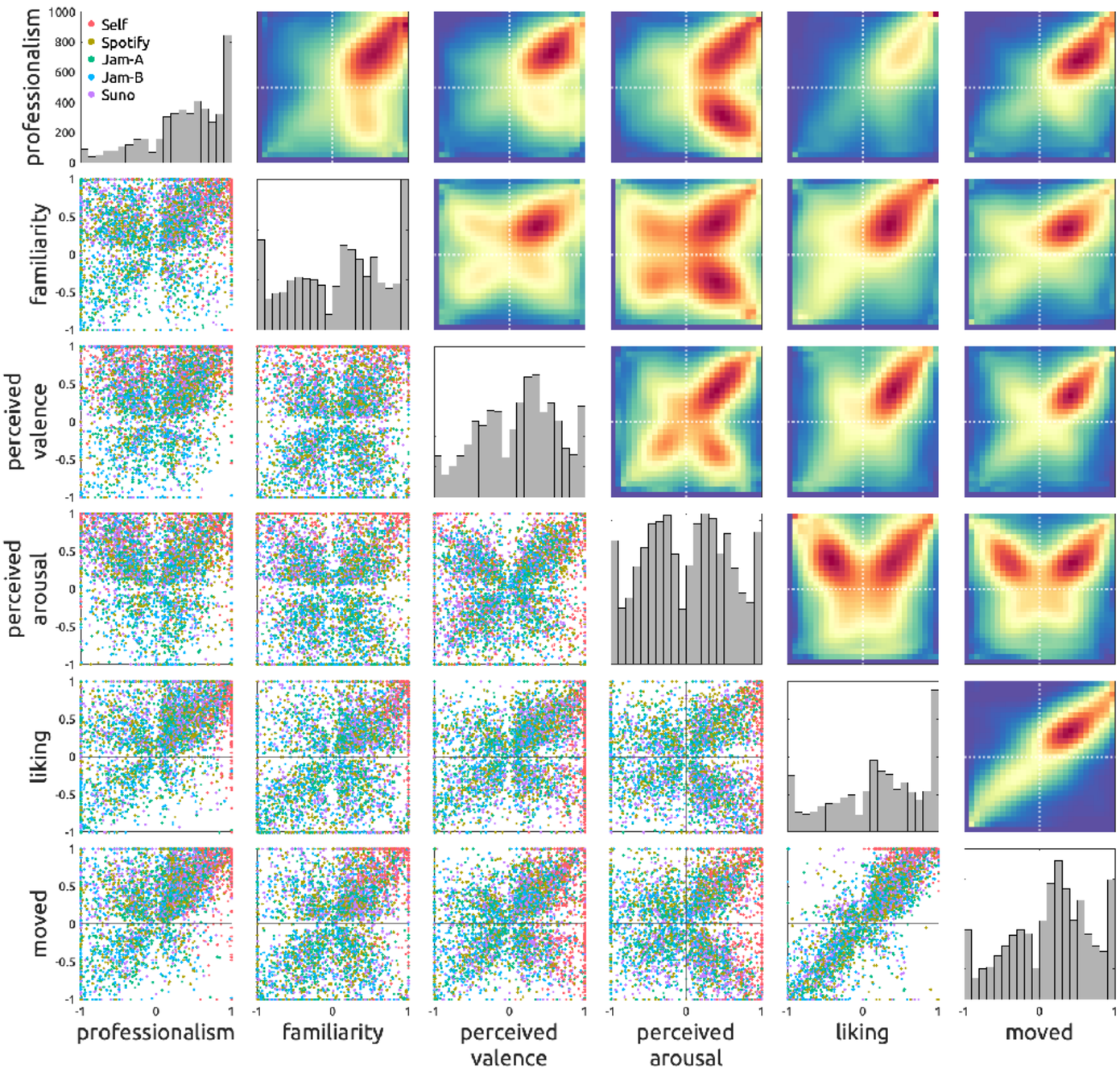


Figure 2: Ridgeline plots (190 participants, 4,600 ratings). Mean kernel bandwidth was 0.233 ± 0.197. White vertical bars denote the 95% confidence intervals of the means. White circles denote medians.



Example MTG-Jamendo tracks

- Most disliked Blues
- Most liked Blues

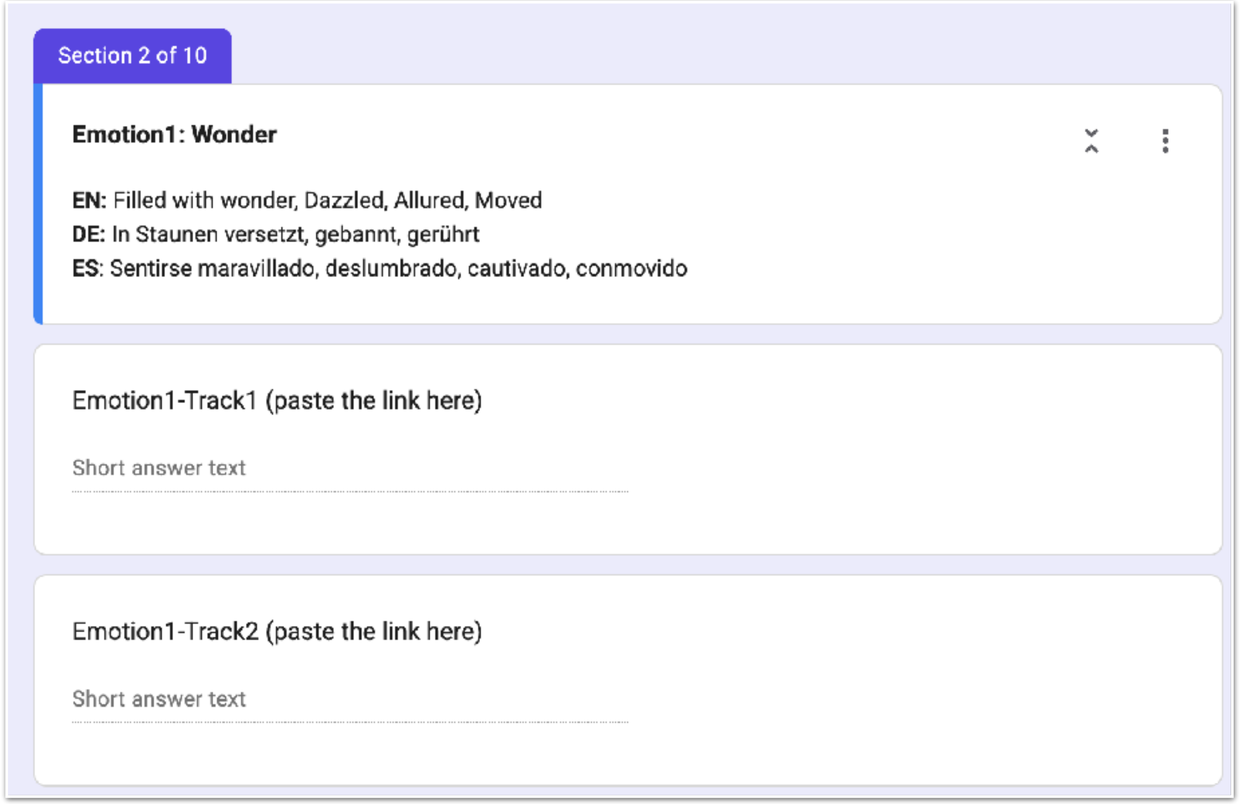


Music selection (Online task)

Participant-selected music

GEMS (Geneva Emotional Music Scale): 9 emotions x 5 tracks

1.	**Wonder** Filled with wonder, Dazzled, Allured, Moved
2.	**Transcendence** Fascinated, Overwhelmed, Feelings of transcendence and spirituality
3.	**Power** Strong, Triumphant, Energetic, Fiery
4.	**Tenderness** Tender, Affectionate, In love, Mellowed
5.	**Nostalgia** Nostalgic, Dreamy, Sentimental, Melancholic
6.	**Peacefulness** Serene, Calm, Soothed, Relaxed
7.	**Joyful Activation** Joyful, Amused, Animated, Bouncy
8.	**Sadness** Sad, Sorrowful
9.	**Tension** Tense, Agitated, Nervous, Irritated

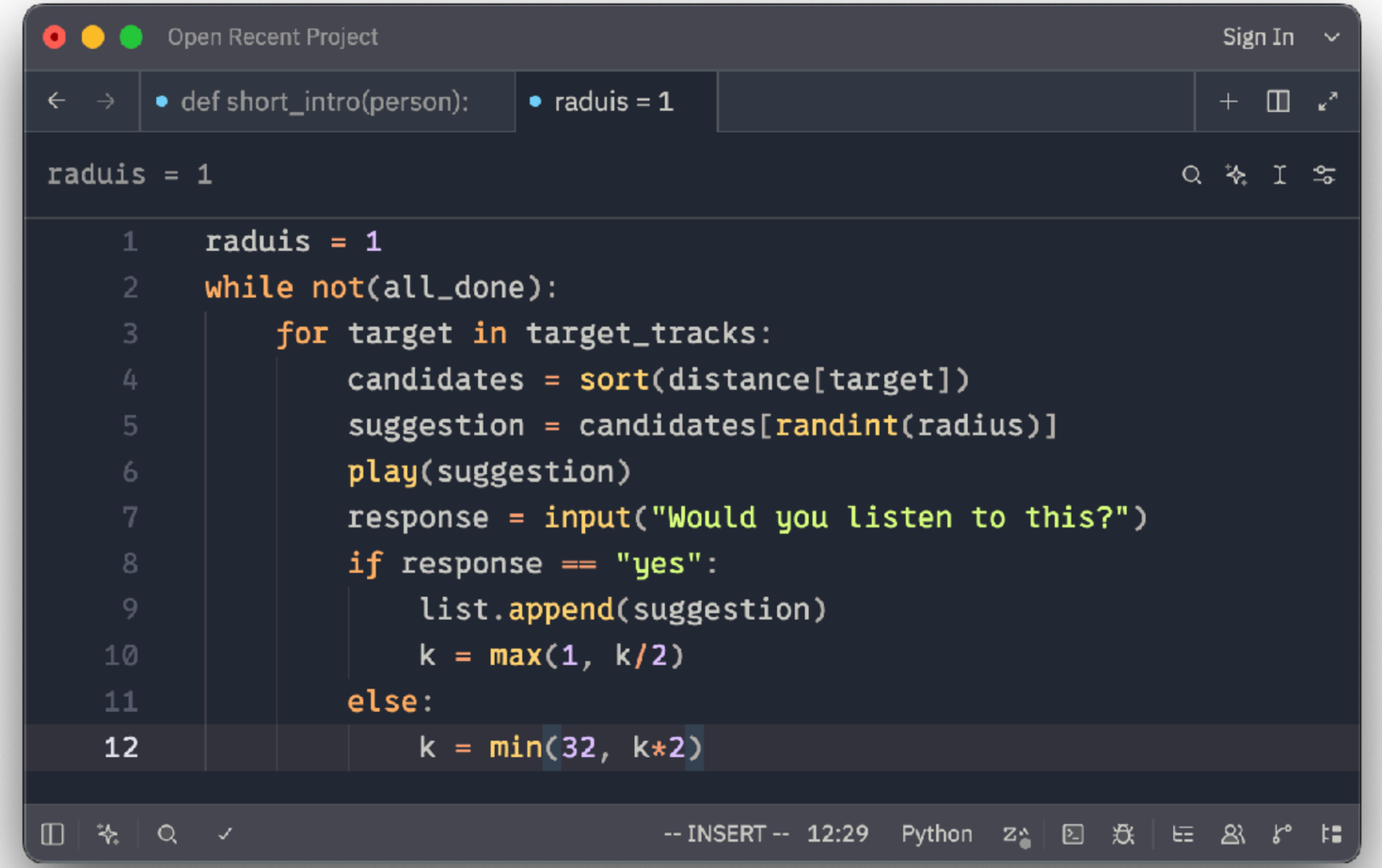


(C) Zenter

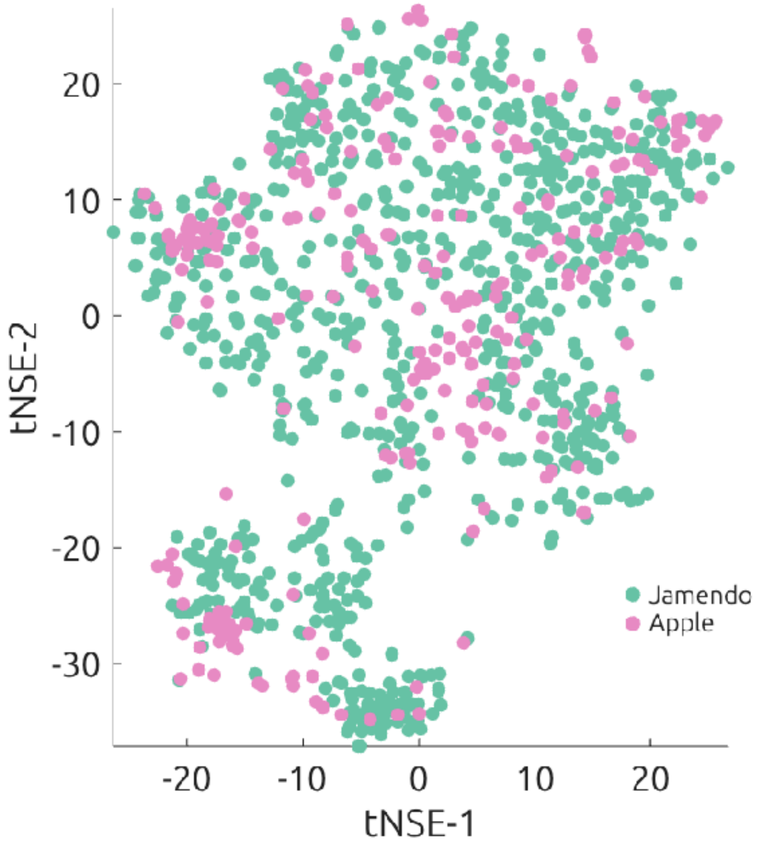
Experimenter-recommended music

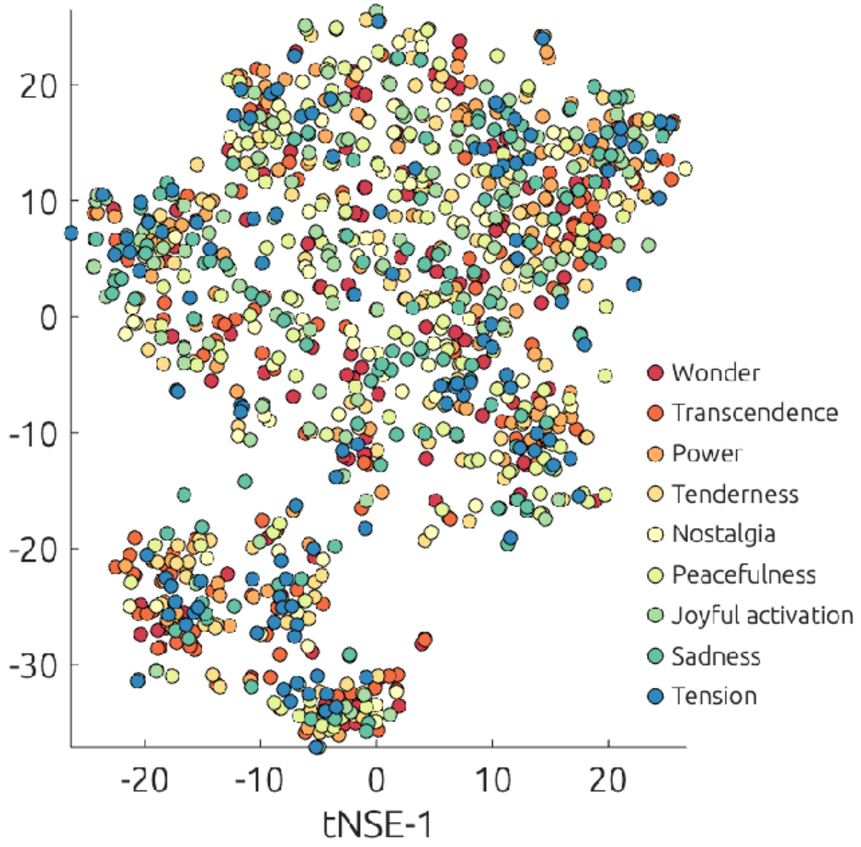
Participants were to accept or reject recommendations.

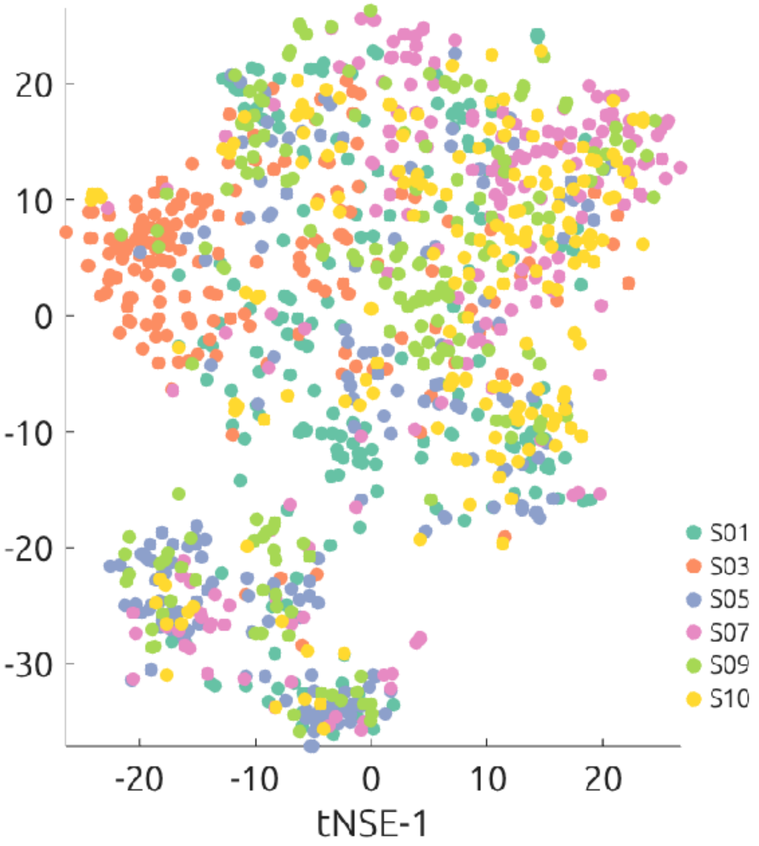
- Sort tracks Jamendo (N=2599) by the similarity to participant-selected tracks
- Distance based on a music transformer model (MAEST)
- For 45 targets, the search was continued until 45*3 tracks were accepted.



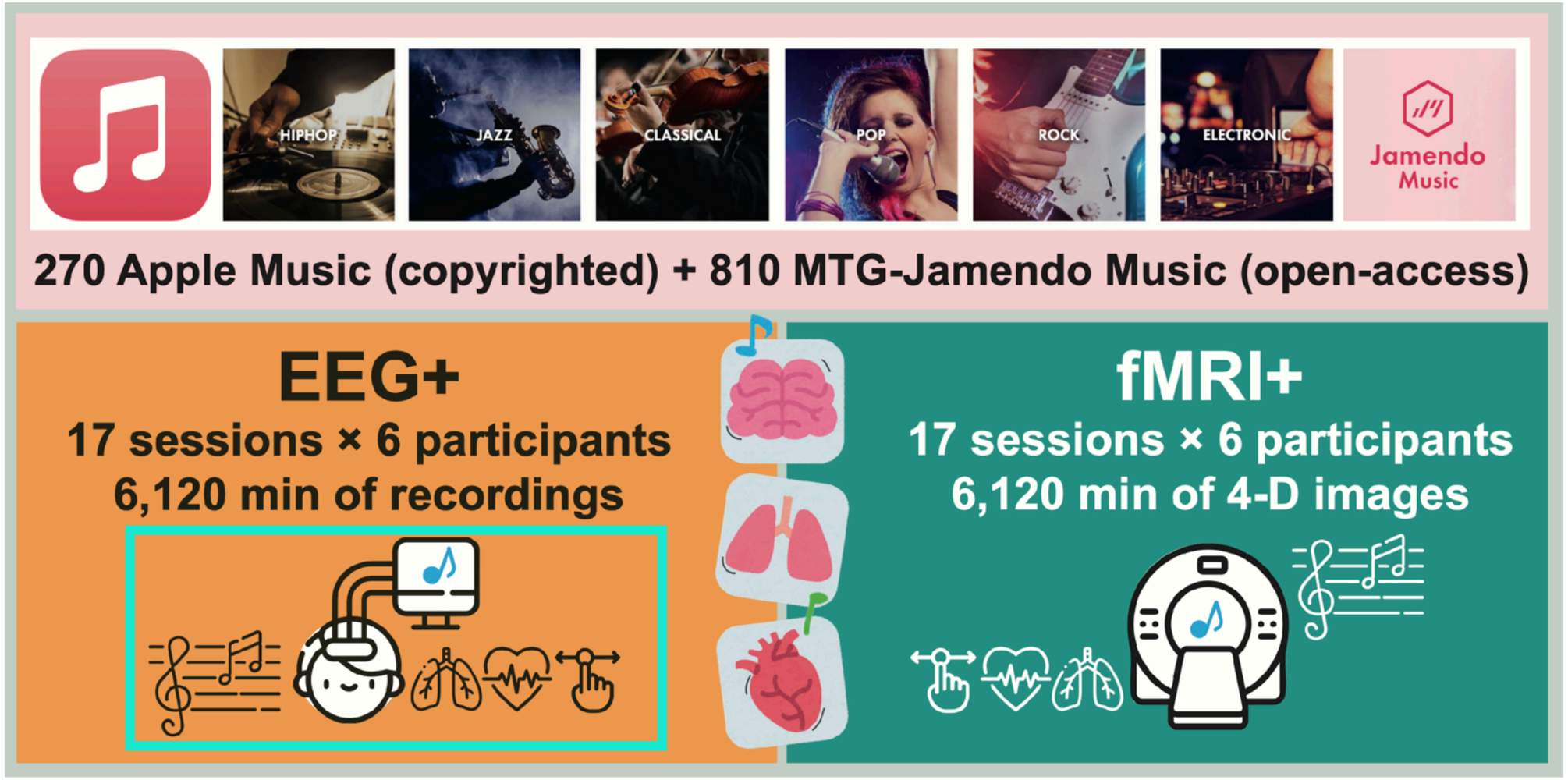
Overview of selected tracks (N=1089)







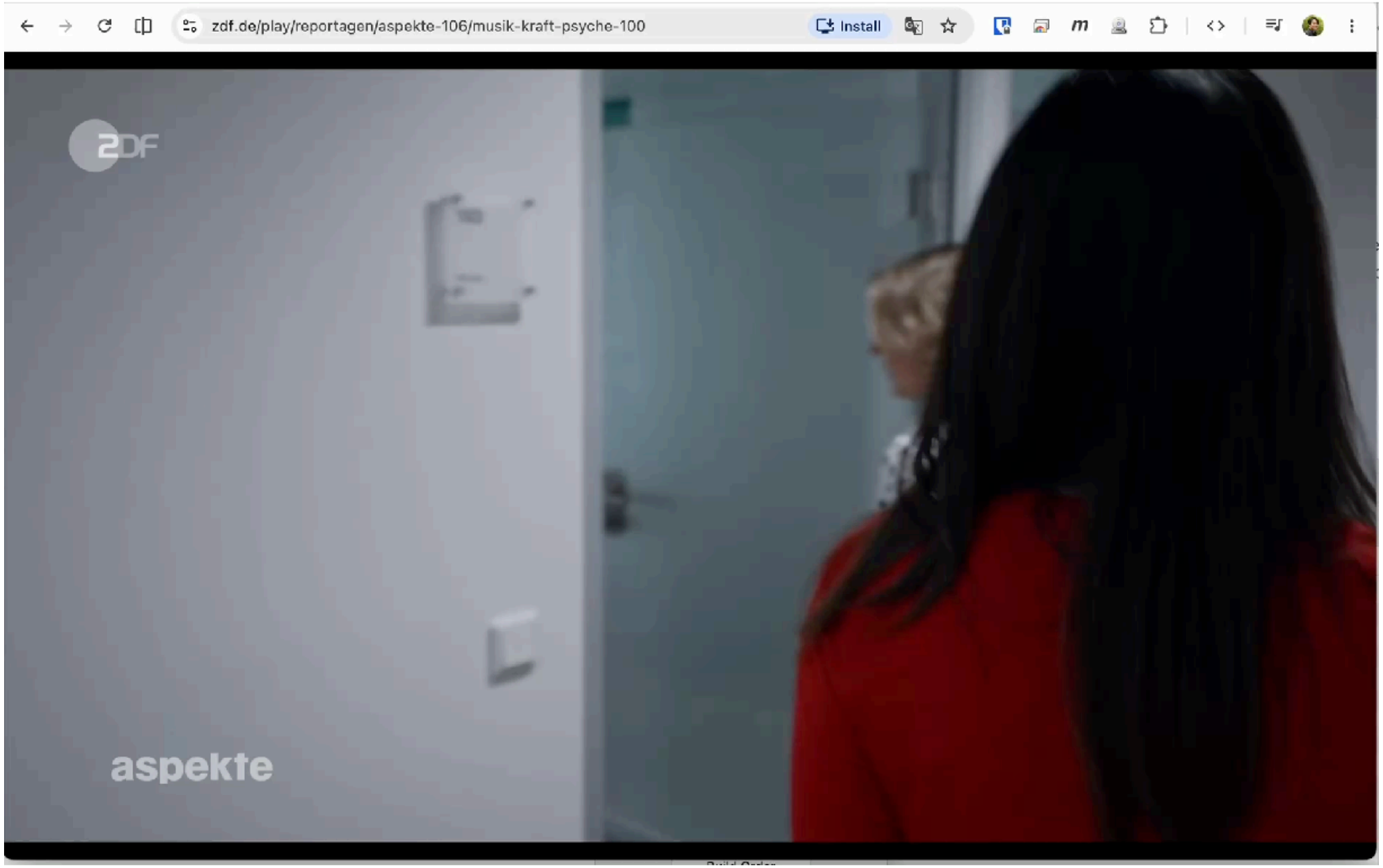
Design and Components



https://flaticon.com

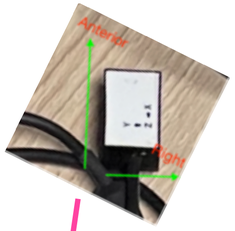
ZDF sketch of ManyMusic-EEG (2026)

https://www.zdf.de/video/reportagen/aspekte-106/musik-kraft-psyche-100



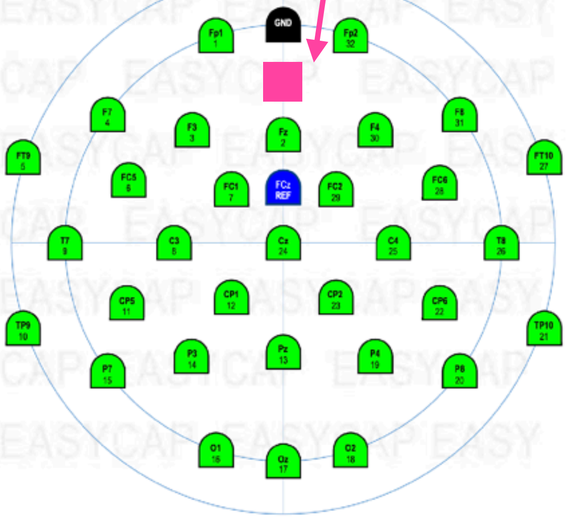
Acquisition setting (EEG+)

32-EEG + 2-EOG + 2-fEMG + SCR + 3-ACC + PULSE + RESP + trackball



Electrode Names and Number Labels

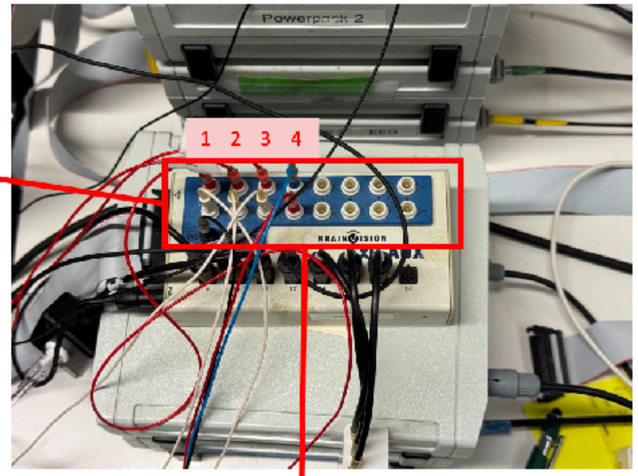
actiCAP

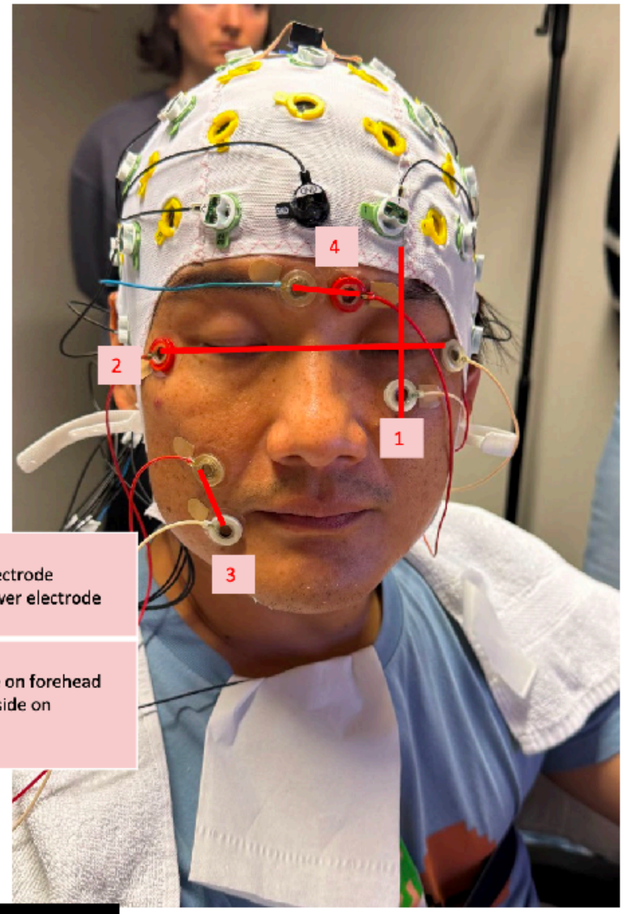


28	FC6	EEG	28	0.1	
29	FC2	EEG	29	0.1	
30	F4	EEG	30		
31	F8	EEG	31		
33	EOGv	BIP	65	10	32
34	EOGh	BIP	66	10	32
35	EMGs	BIP	67	10	32
36	EMGf	BIP	68	10	32
37	STIM	AUX	73	152.6	50
38	SCR	AUX	74	152.6	50
39	RESP	AUX	75	152.6	50
40	PULSE	AUX	76	152.6	50
41	X	AUX	77	152.6	50
42	Y	AUX	78	152.6	50
43	Z	AUX	79	152.6	50

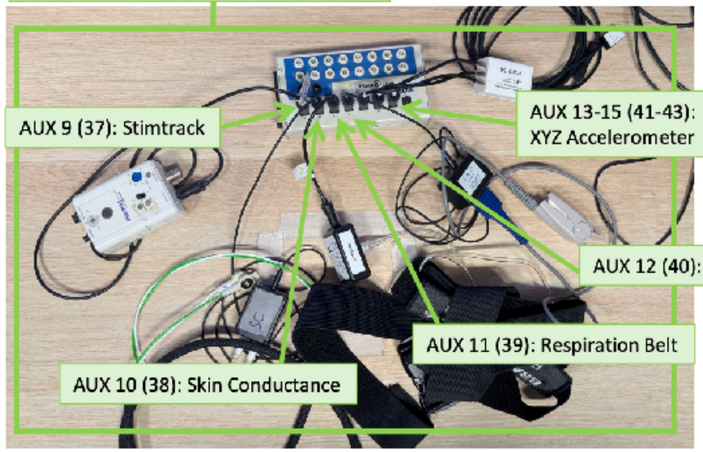
EEG Channels 1-32

Use Electrode Position Fil





AUX Channels 37-43:



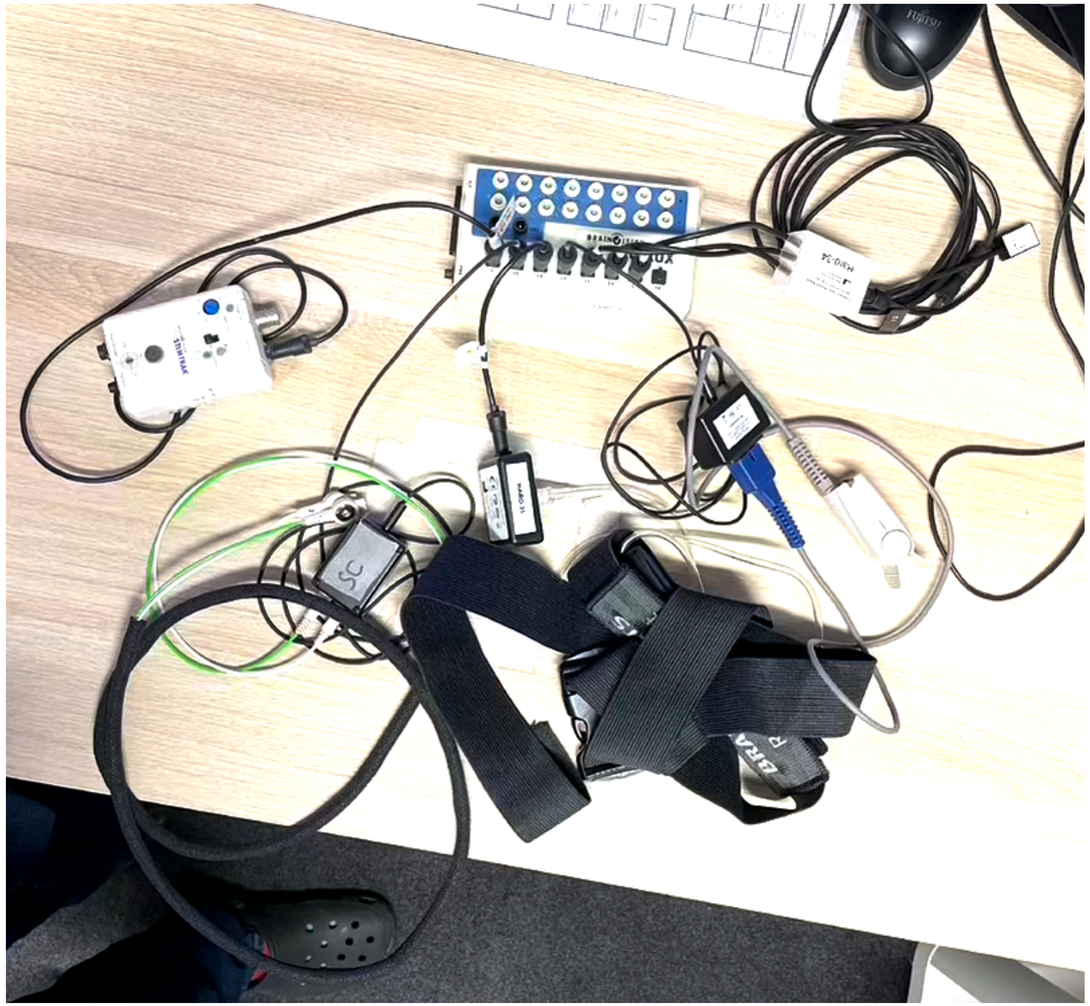
Bipolar Channels 33-36:

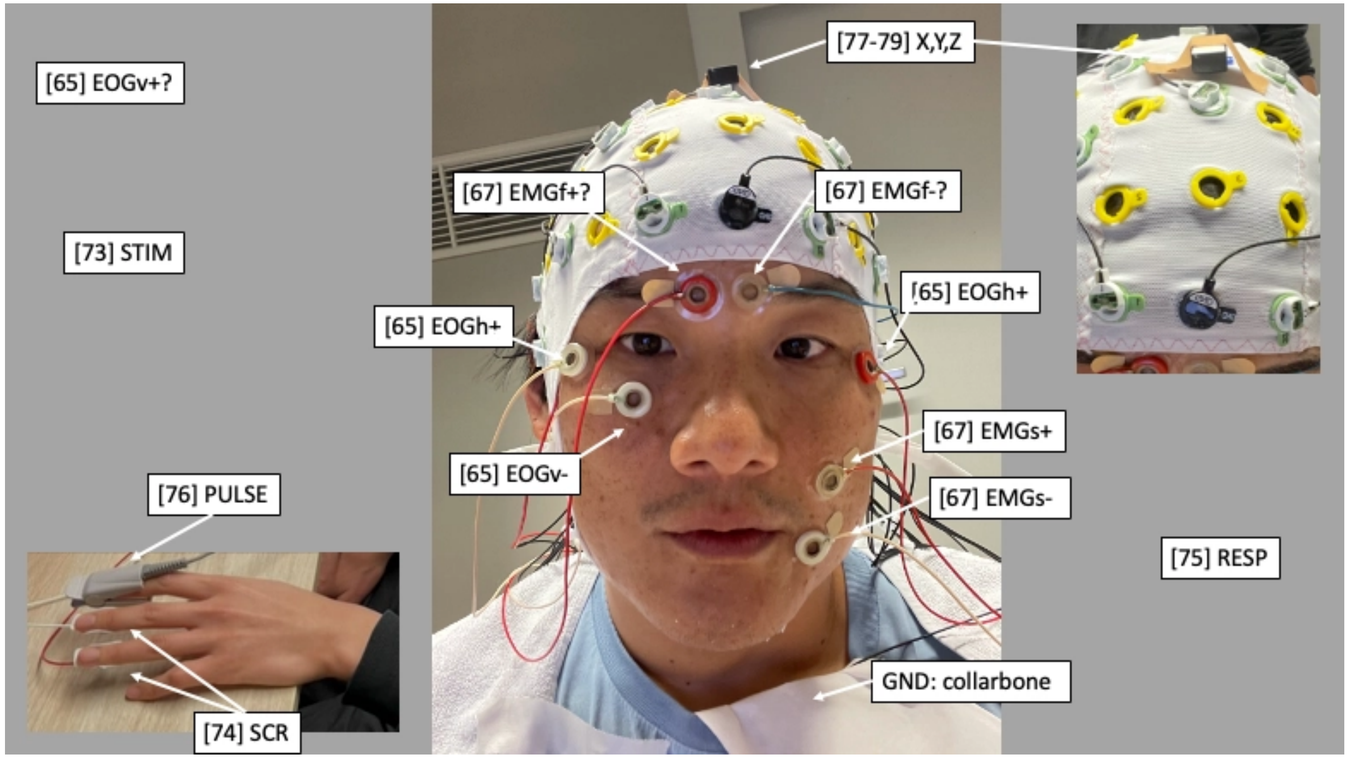
1: EOG vertical (33): Red = top = above Eye White = bottom = below Eye	3: EMG smile (35): Red = top = upper electrode White = bottom = lower electrode
2: EOG horizontal (34): Red = top = right Eye White = bottom = left Eye	4. EMG Frowning (36): Red = top = right side on forehead Blue = bottom = left side on forehead

2023_0360_MUSAFX Setup

Acquisition setting (EEG+)

32-EEG + 2-EOG + 2-fEMG + SCR + 3-ACC + PULSE + RESP + trackball



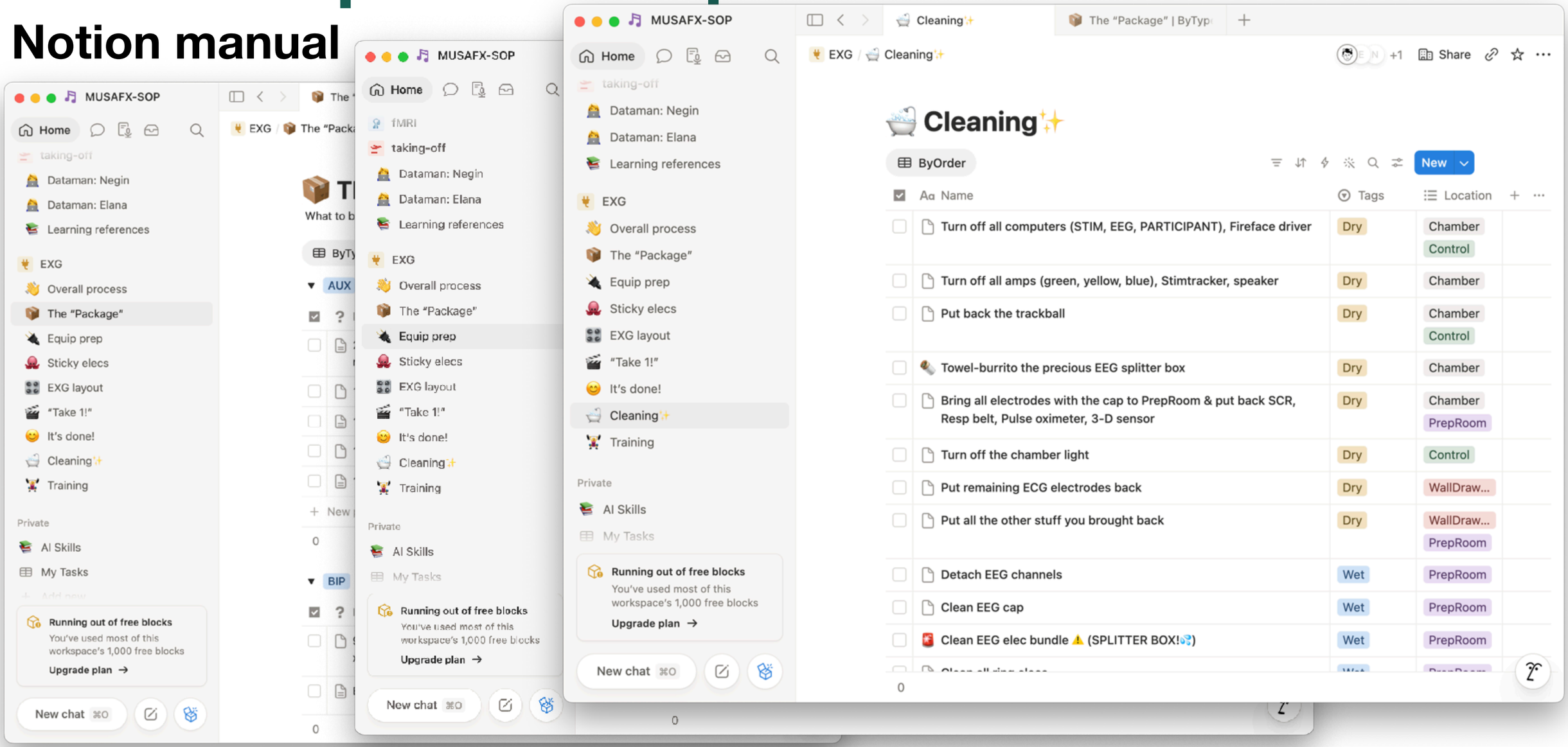




Daily protocol

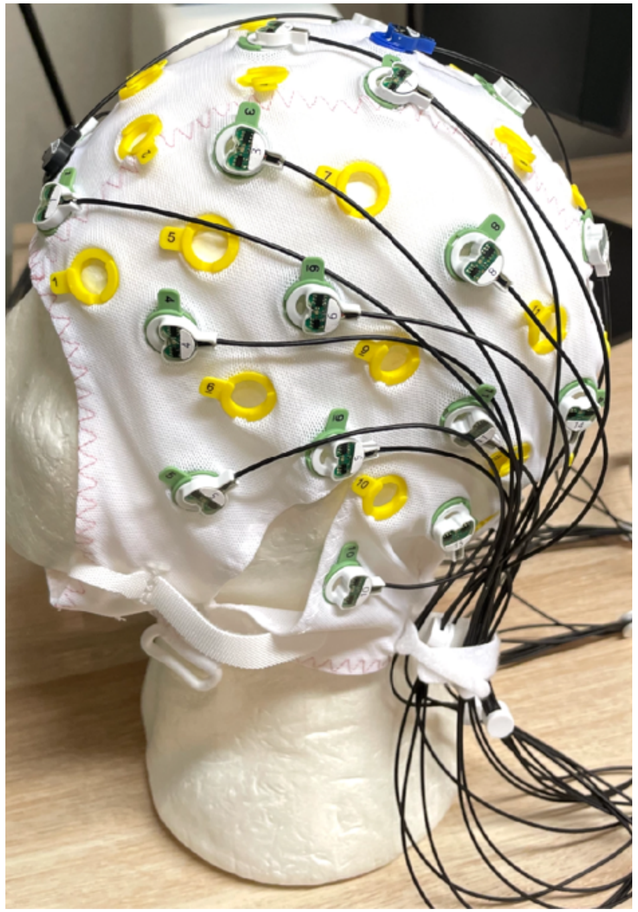
Standard protocol & practice

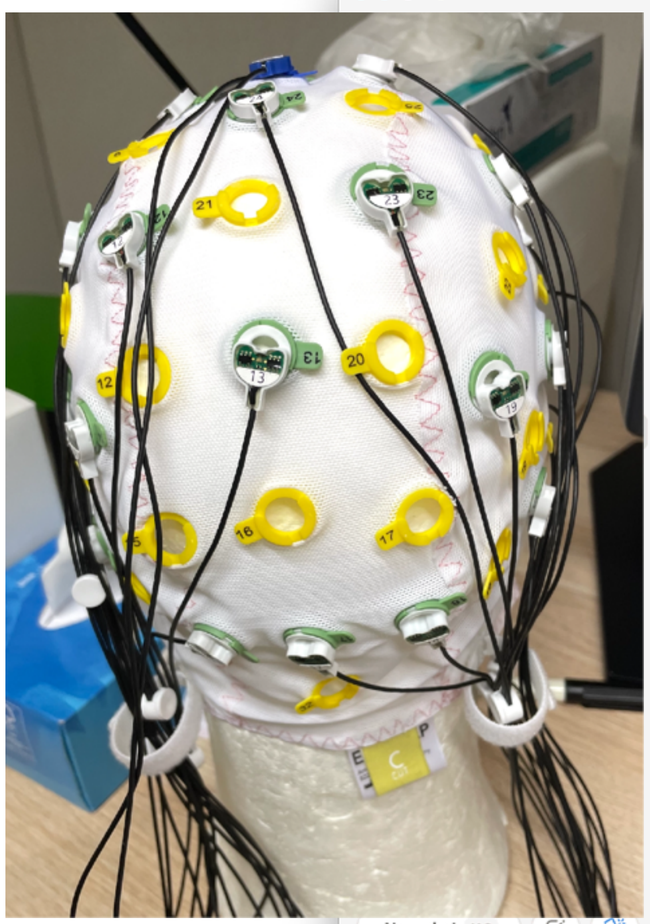
Notion manual



Efforts for consistency

EEG pigtail, labels, the “correct” order, …





MUSAFX-SOP

Sticky elecs

EXG / Sticky elecs

Share

Sticky elecs

The order is a suggestion, not a strict rule, for convenience for the experimenter and participant. Read
Brain Products **Brain Products' actiCAP active Electrodes Walkthrough**

ByOrder

New

Order	ToDo	ChType
1	Respiration belt: chest	AUX
2	Towel over the shoulder	EEG
3	Put on the EEG cap (TIGHTLY while comfortable)	EEG
4	EEG gelling: GND, REF, and Fz w/ impedance check	EEG
5	SCR: electrodes on the left middle & ring fingers	AUX
6	Facial skin prep	EEG
7	EXG-GND: collarbone(right)	BIP
8	EOGs: horizontal, vertical-left(stick and gel with two people)	BIP
9	EMGs: smiling, frowning	BIP
10	PULSE: left index finger	AUX
11	SCR: connect the electrodes	AUX
12	3-D Acc: YELLOW-21 [CPz], cable in the air!	AUX

New page

0 COUNT 12

New chat ⌘O

What participants do

Every week, for ≤3 hours

- Consent form
- PANAS (positive & negative affect schedule): 20-items describing daily mood (e.g., annoyed, bored, excited, …)
- Testing run (15 sec)
- Music listening runs (3–7 min) [x9–13]
- (Cued artefact mapping)
- (Resting with eyes open/closed)



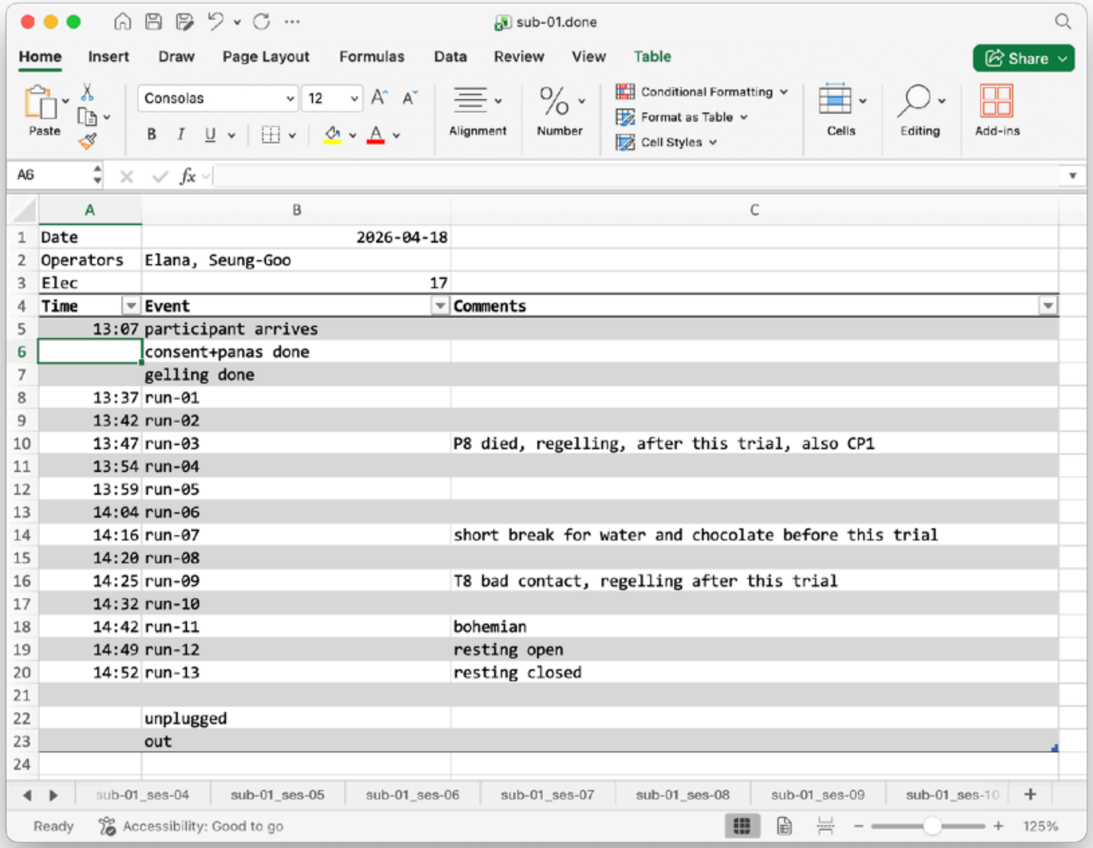
Music listening run

- Participant presses button to start
- Continuous rating of liking while listening to music
- Overall ratings [1–4]: ”Intensity of Overall Emotions", "Absorption to or Immersion in Music", "Enjoyment and Liking", "Goosebumps or Tears”, "Familiarity with the Music Style"
- GEMS ratings [1-4]: Wonder, Transcendence, Power, Tenderness, Nostalgia, Peacefulness, Joyful activation, Sadness, Tension



What we do during acquisition

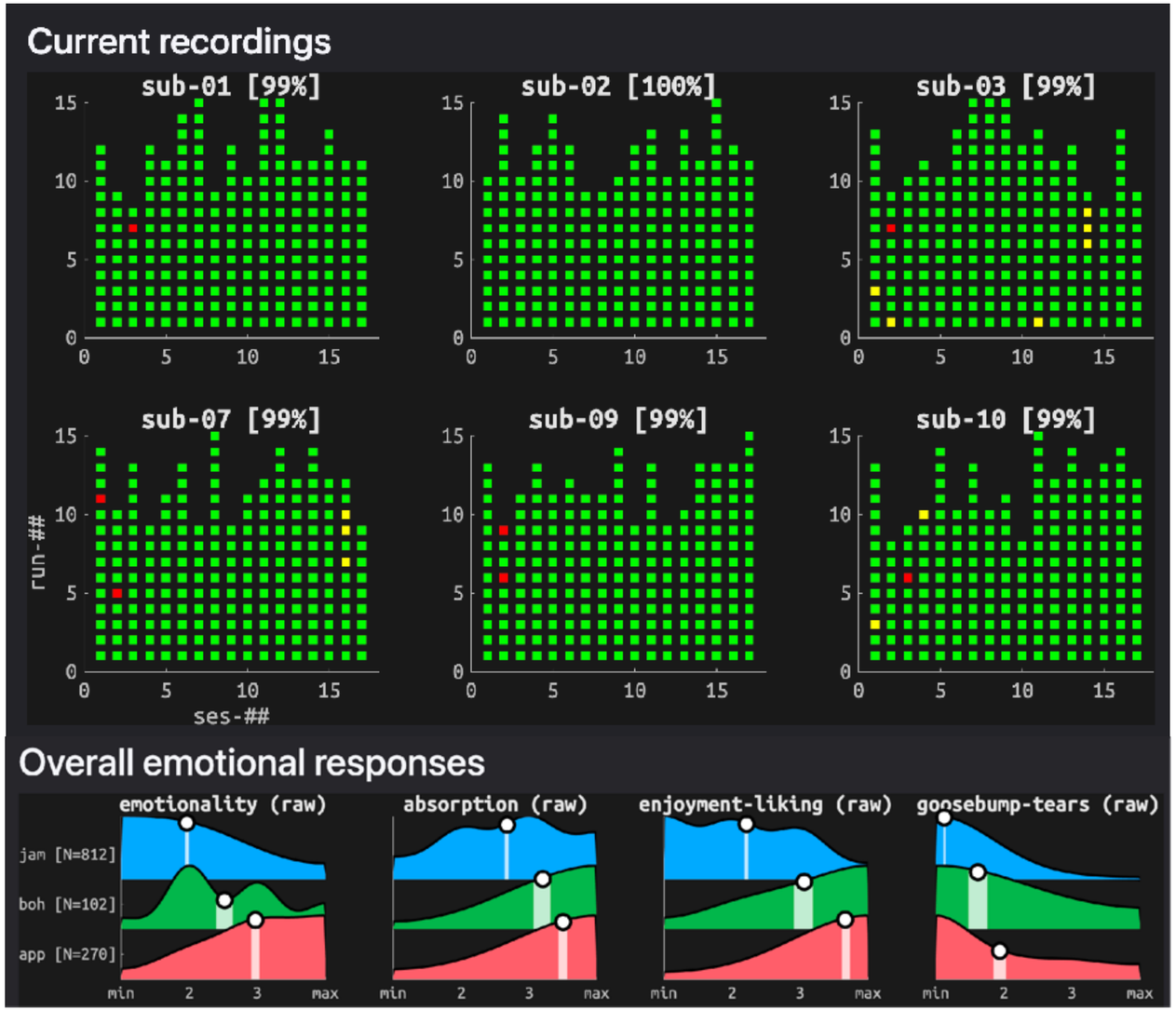
- Make logs for each run.
- See if anything is flatlined or *too noisy*, fix it before the next run.
- Notice if the participants are behaving (not moving too much, not sleeping, …)
- Take short breaks if participants need.

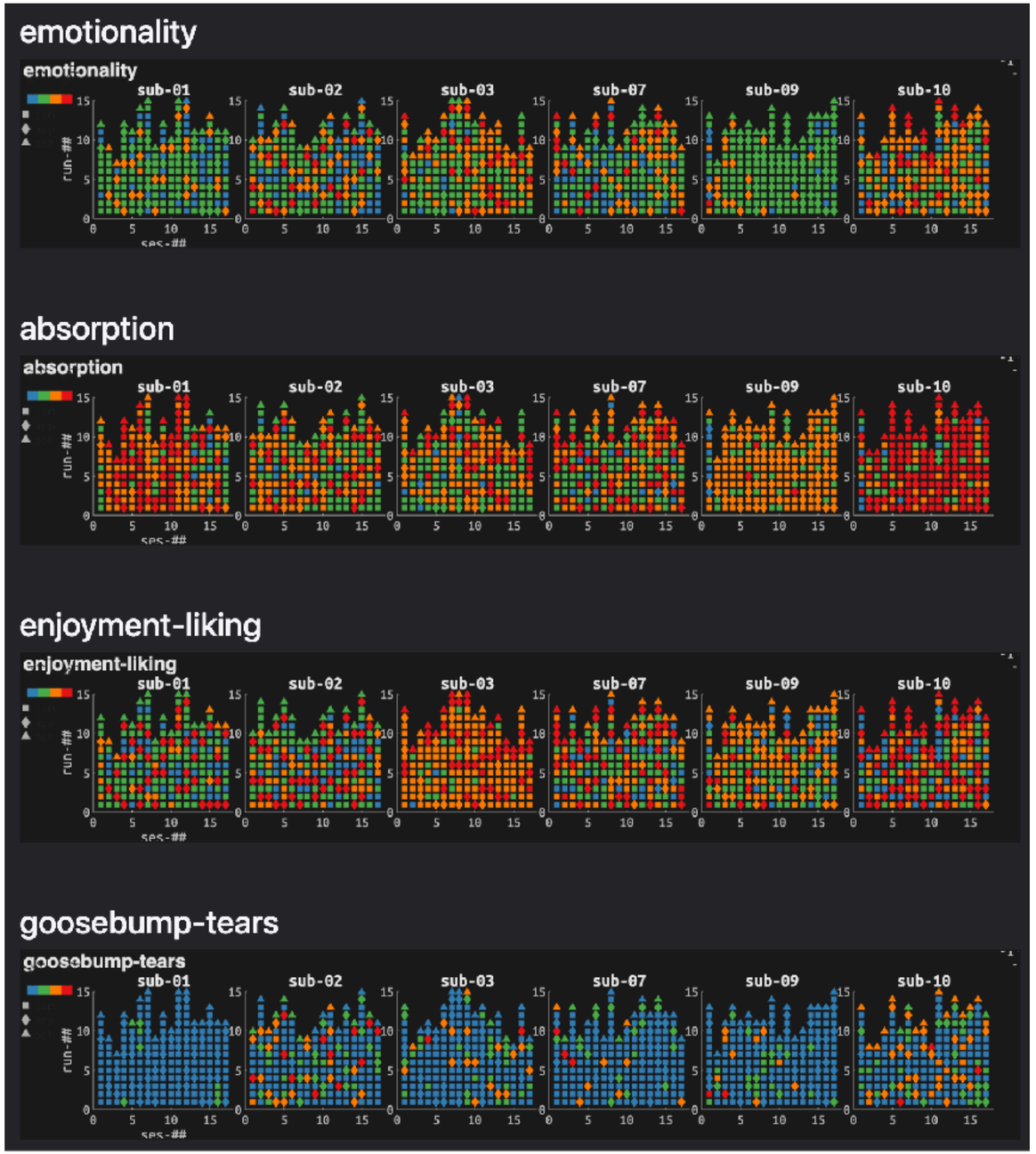


https://manymusic.net/rec/

MAX PLANCK INSTITUTE
FOR EMPIRICAL AESTHETICS

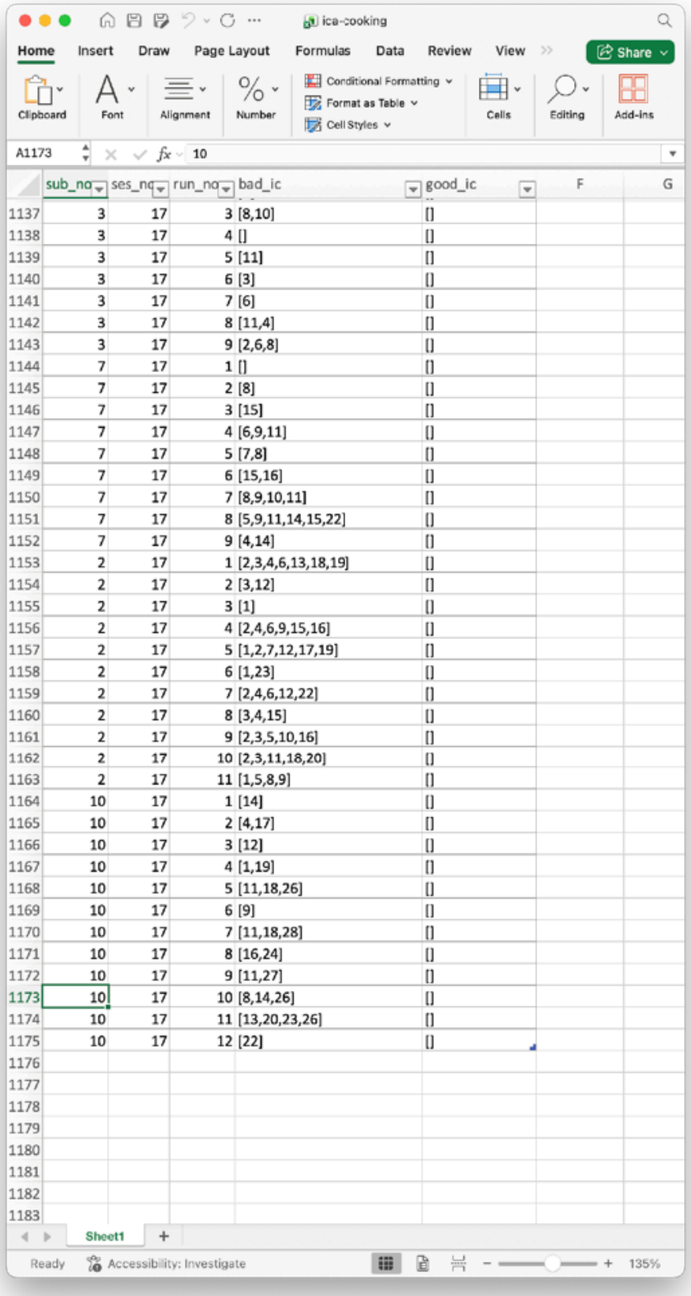
Daily progress monitor





Daily ICA cooking





ica-cooking

Home | Insert | Draw | Page Layout | Formulas | Data | Review | View | Share

Clipboard | Font | Alignment | Number | Conditional Formatting | Format as Table | Cell Styles | Cells | Editing | Add-ins

A1173 | 10

	sub_no	ses_no	run_no	bad_ic	good_ic	F	G
1137	3	17	3	[8,10]	[]		
1138	3	17	4	[]	[]		
1139	3	17	5	[11]	[]		
1140	3	17	6	[3]	[]		
1141	3	17	7	[6]	[]		
1142	3	17	8	[11,4]	[]		
1143	3	17	9	[2,6,8]	[]		
1144	7	17	1	[]	[]		
1145	7	17	2	[8]	[]		
1146	7	17	3	[15]	[]		
1147	7	17	4	[6,9,11]	[]		
1148	7	17	5	[7,8]	[]		
1149	7	17	6	[15,16]	[]		
1150	7	17	7	[8,9,10,11]	[]		
1151	7	17	8	[5,9,11,14,15,22]	[]		
1152	7	17	9	[4,14]	[]		
1153	2	17	1	[2,3,4,6,13,18,19]	[]		
1154	2	17	2	[3,12]	[]		
1155	2	17	3	[1]	[]		
1156	2	17	4	[2,4,6,9,15,16]	[]		
1157	2	17	5	[1,2,7,12,17,19]	[]		
1158	2	17	6	[1,23]	[]		
1159	2	17	7	[2,4,6,12,22]	[]		
1160	2	17	8	[3,4,15]	[]		
1161	2	17	9	[2,3,5,10,16]	[]		
1162	2	17	10	[2,3,11,18,20]	[]		
1163	2	17	11	[1,5,8,9]	[]		
1164	10	17	1	[14]	[]		
1165	10	17	2	[4,17]	[]		
1166	10	17	3	[12]	[]		
1167	10	17	4	[1,19]	[]		
1168	10	17	5	[11,18,26]	[]		
1169	10	17	6	[9]	[]		
1170	10	17	7	[11,18,28]	[]		
1171	10	17	8	[16,24]	[]		
1172	10	17	9	[11,27]	[]		
1173	10	17	10	[8,14,26]	[]		
1174	10	17	11	[13,20,23,26]	[]		
1175	10	17	12	[22]	[]		
1176							
1177							
1178							
1179							
1180							
1181							
1182							
1183							

Sheet1 +

Ready | Accessibility: Investigate | 135%



Trouble shooting

Repair notes

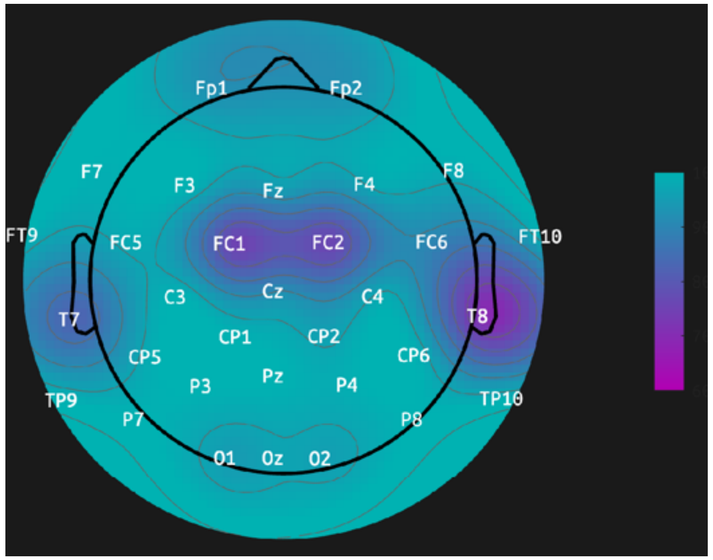
https://manymusic.net/rec/repairs.html

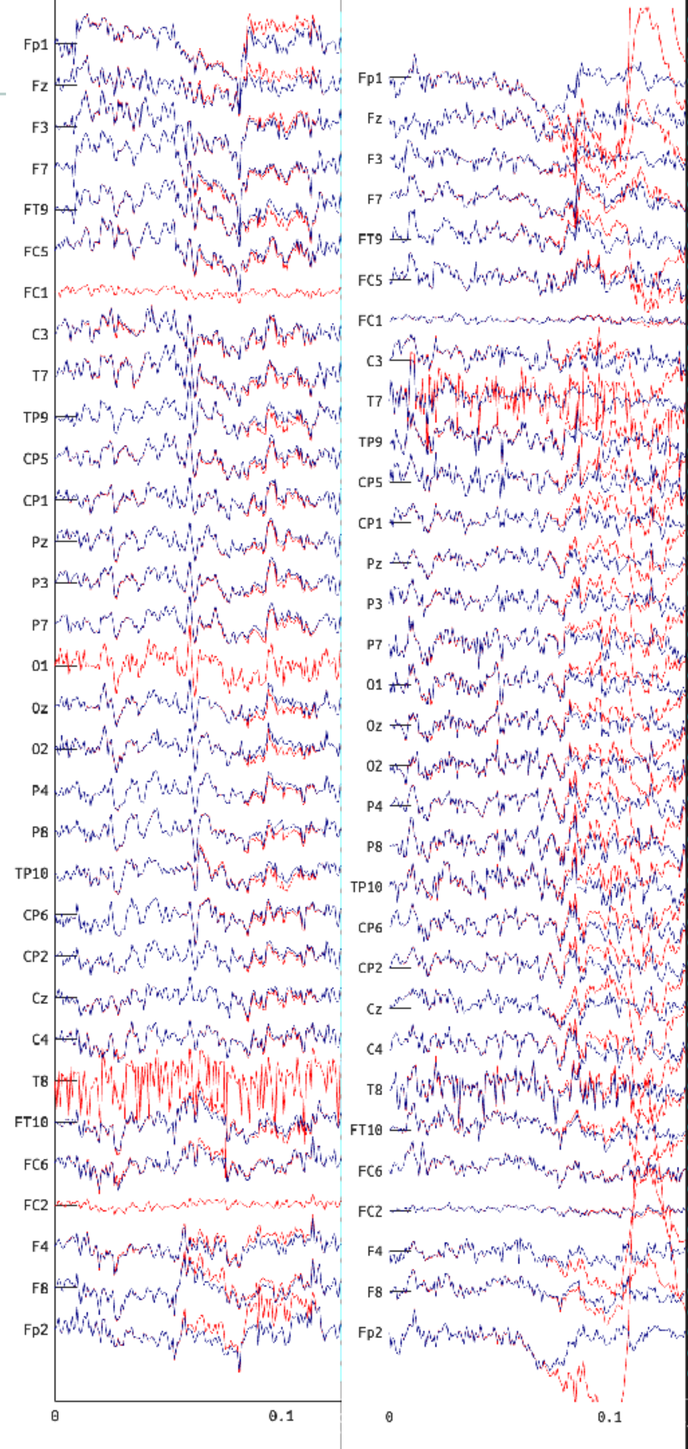
- Problems with the audio source
  - 3 tracks were in different sampling rates x my bug in the code
  - 7x2 tracks were identical but with different track-IDs
- NIRCMD memory leakage
- Hot temperature in the EEG chamber (~April 2026)

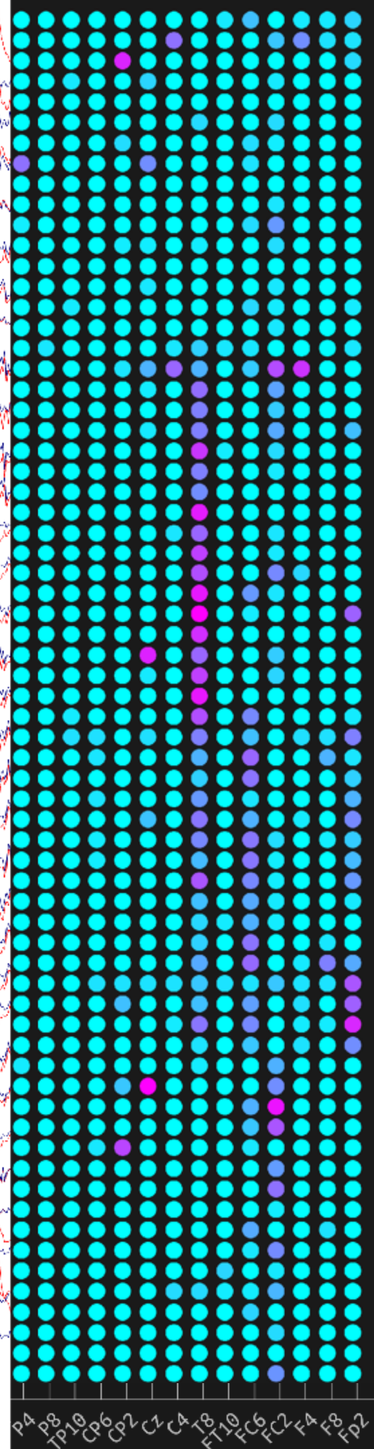
Repair notes

https://manymusic.net/rec/repairs.html

- Noisy FC1 & FC2 for two subjects
  - Smaller caps, monitoring & gelling







Any questions so far?

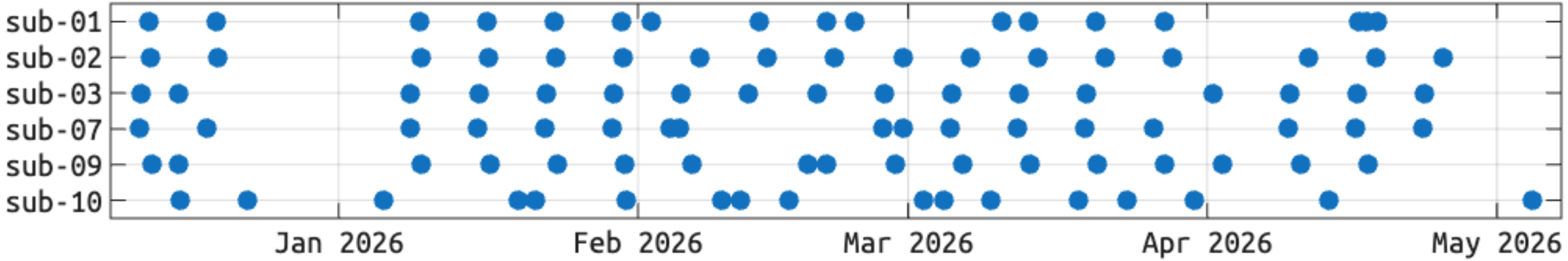
Workshop plan

Building a 1,000-Song Naturalistic EEG Dataset

Part	From	To	Contents	Dur
1	09:00	09:45	**Introduction**	45m
2	10:00	10:45	**Analysis**	45m
3	11:00	11:45	**Design**	45m
4	12:00	12:45	**Management**	45m

102 sessions over 5 months

Too many or too few?

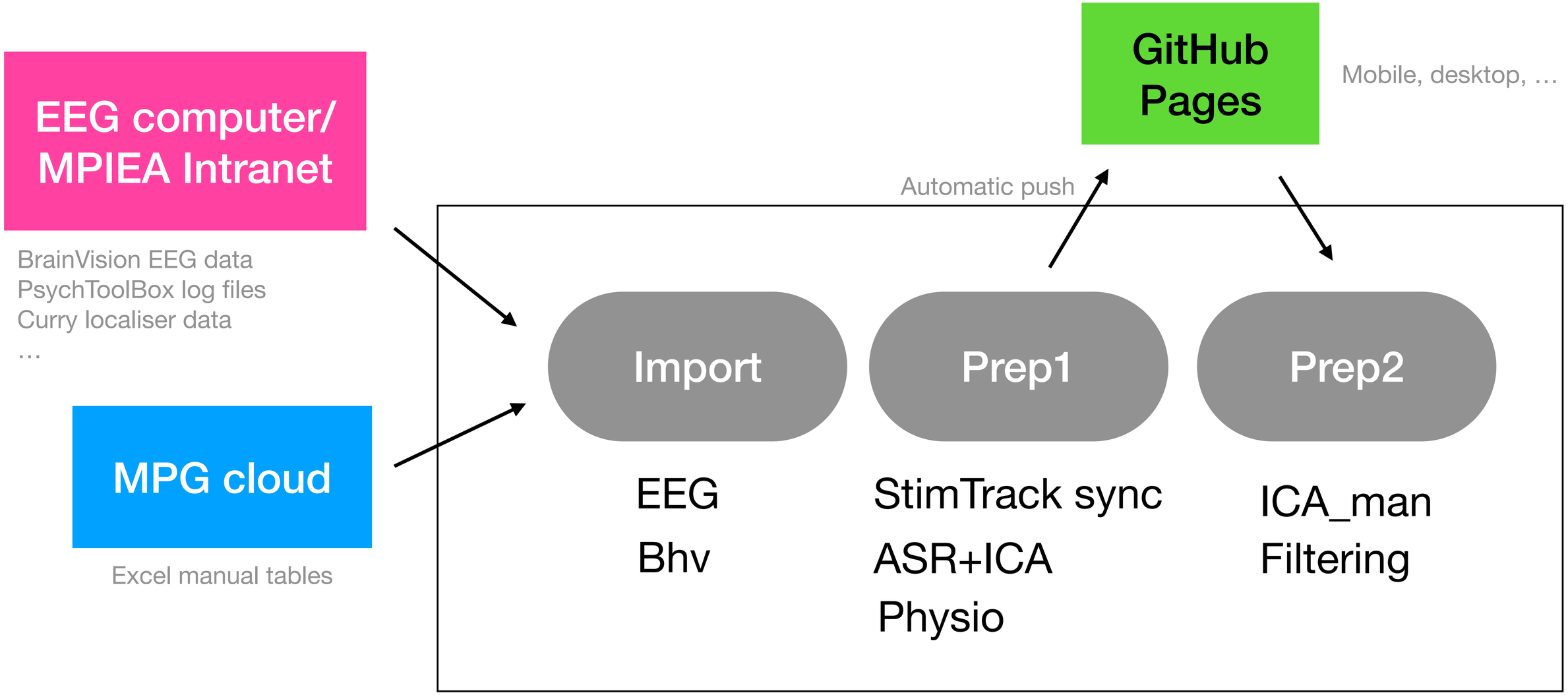


Daily monitoring pipeline

Development process



Daily monitoring pipeline



Additional tasks for data quality check

- Cued-artefacts
- Resting state
- Same song at the end of all sessions (Bohemian rhapsody)



Cued-artefacts

Cued-artefacts mapping

Trackball tracing

- Q: What are the trackball-control-related EXG artefacts?
- https://manymusic.net/rec/pages/tracing/parent.html

Cued-artefacts mapping

Do <TASK> on beeps

- Q: What are the influence of artefacts to EEG?
- Frowning, jaw clenching, nodding, heavy breathing, foot tapping
- https://manymusic.net/rec/pages/artefacts/artefacts.html



Resting

Resting-state data

- Q: how different is music listening from resting (w/ eyes open/closed)? (e.g., neural oscillation; or micro-state)
- https://manymusic.net/rec/pages/resting/resting-microstates.html
- Kalburgri+ (2024) MICROSTATELAB. https://doi.org/10.1007/s10548-023-01003-5

Bohemian rhapsody for reliability check

- Q: Are the multiple EEG+ sessions consistent across time?
- https://manymusic.net/rec/pages/bohemian/bohemian.html

DIY hands-on

Exercises: does daily mood affect musical emotion?

CSV table: https://s.gwdg.de/WbbzvP

- RQ1: Are there any associations between PANAS and aggregated ratings (e.g., GEMS, liking, …)?
- RQ2: Are there any associations between PANAS and average physiology during the session?
- RQ3: Are there any seasonal influences on PANAS?
- RQ*: or any other RQs?
- <Use any stat model in any language (or any agent)>

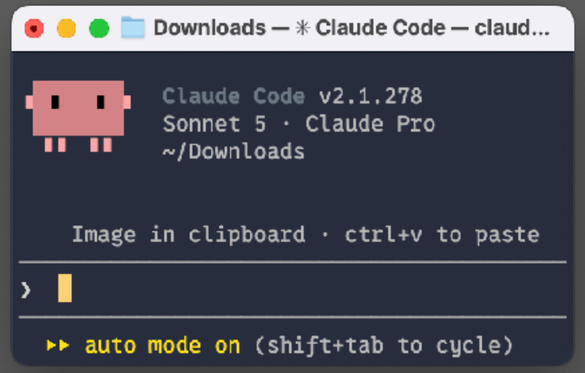




Table cookbook

data.csv

- DEMO: subId, sesId, runId, runDate
- PANAS: panas_* [interested, distressed, excited, upset, strong, guilty, scared, hostile, enthusiastic, proud, irritable, alert, ashamed, inspired nervous, determined, attentive, jittery, active, afraid] 1=min, 5=max
- GEMS: gems_* [Wonder, Transcendence, Power, Tenderness, Nostalgia, Peacefulness, Joyful activation, Sadness, Tension]; 1=min, 4=max
- RATINGS: rating_* [Intensity, Absorption, Liking, Goosebumps] 1=min, 4=max
- PHYSIO: heartRateMean_bpm, heartRateStd_bpm, respRateMean_bpm, respRateStd_bpm

MAX PLANCK INSTITUTE
FOR EMPIRICAL AESTHETICS

Thank you for your attention!